GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:11:08 ; Search time 1346.95 Seconds
(without alignments)
10435.915 Million cell updates/sec

Title: US-09-677-374-3
Perfect score: 483
Sequence: 1 atgcgtggttgcctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	295	61.1	489	6	AX252413	AX252413 Sequence
	2	295	61.1	4983	1	AF184152	AF184152 Piscirick
	3	46.6	9.6	537	1	RIR17KGCA	D16515 Rickettsia
	4	45	9.3	237523	1	RPXX04	AJ235273 Rickettsi
	5	42.8	8.9	203050	1	AL646071	AL646071 Ralstonia
	6	41.6	8.6	448	1	RSU76907	U76907 Rickettsia
	7	41	8.5	198050	1	AL646061	AL646061 Ralstonia
	8	40.8	8.4	213050	1	AL646079	AL646079 Ralstonia
	9	40.2	8.3	539	1	RIRANT17KA	M28480 R.conorii 1
	10	40.2	8.3	539	1	RIRANT17KD	M28479 R.rickettsi
	11	40.2	8.3	10127	1	AE008675	AE008675 Rickettsi
	12	39.6	8.2	13946	1	AE004537	AE004537 Pseudomon
	13	39.4	8.2	106702	8	ATF7K2	AL033545 Arabidops
	14	39.4	8.2	199577	8	ATCHRIV57	AL161557 Arabidops
c	15	39	8.1	1070	3	DMU60844	U60844 Drosophila
c	16	39	8.1	1070	3	DMU60845	U60845 Drosophila
c	17	39	8.1	1070	3	DMU60847	U60847 Drosophila
c	18	39	8.1	1070	3	DMU60850	U60850 Drosophila
c	19	38.8	8.0	262428	2	AC096326	AC096326 Rattus no
	20	38.6	8.0	535	1	RIRANT17KB	M28481 R.typhi 17K
	21	38.6	8.0	620	1	RIRANTRR	M16486 R.rickettsi
c	22	38.4	8.0	125020	9	AF429315	AF429315 Homo sapi
c	23	38	7.9	2820	1	SVI17269	Y17269 Streptomyce
	24	38	7.9	167184	2	AP005650	AP005650 Oryza sat
	25	37.8	7.8	13123	1	AE003872	AE003872 Xylella f
	26	37.8	7.8	61958	2	AC121467	AC121467 Rattus no
	27	37.6	7.8	10887	1	AE011779	AE011779 Xanthomon
c	28	37.4	7.7	804	3	DSU60862	U60862 Drosophila
c	29	37.4	7.7	804	3	DSU60864	U60864 Drosophila
c	30	37.4	7.7	804	3	DSU60866	U60866 Drosophila
c	31	37.4	7.7	804	3	DSU60868	U60868 Drosophila
c	32	37.4	7.7	804	3	DSU60869	U60869 Drosophila
c	33	37.4	7.7	1070	3	DMU60846	U60846 Drosophila
c	34	37.4	7.7	1070	3	DMU60848	U60848 Drosophila
	35	37.4	7.7	1140	1	AB046994	AB046994 Streptomy
	36	37.4	7.7	6629	1	AB072568	AB072568 Streptomy
	37	37.4	7.7	10029	1	AE008032	AE008032 Agrobacte
	38	37.4	7.7	10606	1	AE009066	AE009066 Agrobacte
c	39	37.4	7.7	187647	2	AC106349	AC106349 Rattus no
	40	37.2	7.7	416	1	AF031534	AF031534 Rickettsi
	41	37	7.7	9811	14	AF449714	AF449714 Cercopith
c	42	37	7.7	13348	1	AE005945	AE005945 Caulobact
	43	36.8	7.6	492	1	RIRTRAPRO	M74042 R.australis
c	44	36.8	7.6	1073	3	DSU60861	U60861 Drosophila
c	45	36.8	7.6	6526	1	SCBAC14E8	AL590435 Streptomy

ALIGNMENTS

RESULT 1
AX252413
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168865.
ACCESSION AX252413
VERSION AX252413.1 GI:15985721
KEYWORDS .
SOURCE Piscirickettsia salmonis.
ORGANISM Piscirickettsia salmonis
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
Piscirickettsia.
REFERENCE 1 (bases 1 to 489)
AUTHORS Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and
Burzio,L.
TITLE Fish vaccine against piscirickettsia salmonis

```
  JOURNAL   Patent: WO 0168865-A 5 20-SEP-2001;
            Aqua Health (Europe) Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Piscirickettsia salmonis"
                     /db_xref="taxon:1238"
BASE COUNT      139 a     79 c    144 g    127 t
ORIGIN

  Query Match             61.1%;  Score 295;  DB 6;  Length 489;
  Best Local Similarity   76.0%;  Pred. No. 8.4e-62;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64
Db       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy      65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124
Db      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy     125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184
Db     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy     185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244
Db     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy     245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304
Db     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy     305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364
Db     308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy     365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424
Db     368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy     425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483
Db     428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 2
AF184152
LOCUS       AF184152               4983 bp    DNA     linear   BCT 01-OCT-2000
DEFINITION  Piscirickettsia salmonis alanine racemase (alr) gene, partial cds;
            and BAX (bax), DNA repair enzyme RadA (radA), 17 kDa antigen
            (ospA), and transposase (tnpA) genes, complete cds.
ACCESSION   AF184152
VERSION     AF184152.1  GI:10441343
KEYWORDS    .
SOURCE      Piscirickettsia salmonis.
  ORGANISM  Piscirickettsia salmonis
            Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
            Piscirickettsia.
REFERENCE   1  (bases 1 to 4983)
  AUTHORS   Kuzyk,M.A., Burian,J., Thornton,J.C. and Kay,W.W.
  TITLE     Identification of a genus-common Rickettsial surface antigen in the
            salmonid pathogen Piscirickettsia salmonis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4983)
  AUTHORS   Kuzyk,M.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Biochemistry & Microbiology, University of
            Victoria, Petch Building, PO Box 3055, Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .4983
                     /organism="Piscirickettsia salmonis"
                     /strain="LF-89"
                     /db_xref="taxon:1238"
                     /clone="pB12"
     source          3560. .4849
                     /organism="Piscirickettsia salmonis"
                     /strain="LF-89"
                     /db_xref="taxon:1238"
                     /clone="pB12"
                     /insertion_seq="insertion sequence-like element"
     gene            <1. .501
                     /gene="alr"
     CDS             <1. .501
                     /gene="alr"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="alanine racemase"
                     /protein_id="AAG17001.1"
                     /db_xref="GI:10441347"
                     /translation="NSAAIFNFSYERYDWVRPGIMLYGISPFADKNGVDLELQPVMHV
                     VSRLISVKQLRQGESVGYGATWQCPEDMQVGILSLGYGDGYPRLAASGTPFLVRGQRC
                     ALIGRVSMDMIAIDLRRCPDAGVGEAVTVWGQDLPVEEIARHVGTIAYELVCNMPLRA
                     PYIWQE"
     gene            complement(526. .1257)
                     /gene="bax"
     CDS             complement(526. .1257)
                     /gene="bax"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="BAX"
                     /protein_id="AAG17002.1"
                     /db_xref="GI:10441348"
                     /translation="MLSLAVITTAYLAQQPLFSAPTSKTQINTAYSQDIQSIKQEFII
                     KITQSENTINQIILQKRQHILSLIQKKSLSKQEITQLQALAKQYKLNNLHFKHPQDFS
                     LLLKRVDIIPNALISAQAINESNWGRSRFAVEGNNFFGMRCHYPGCGIIPKARPANNH
                     WEVANYSSMTASVRAYIHTLNTHNAYQALRDLRAHMRANHQDVSAFKLAEGLTAYSIK
                     GTKYVELIQTIITRYLSKESTSQIP"
     gene            1373. .2743
                     /gene="radA"
                     /evidence=experimental
     CDS             1373. .2743
                     /gene="radA"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA repair enzyme RadA"
                     /protein_id="AAG16998.1"
                     /db_xref="GI:10441344"
                     /translation="MSKQKNQYVCSDCGGIAKKWLGQCPHCQQWNSLSEVKEVLPNRP
                     GRSQRFAGFAGIDAPKVKCLSEITPEQISRQPVGICEFDRVLGGGLVHGAVILLGGDP
                     GIGKSTLLLQTSVNCTQFGKVLYVTGEESLEQVTLRSKRLGLSQDVDLRLLAETQVER
                     ILKAAEIEQPKVLIVDSIQTIFTESLQSAPGGVAQVRESAAILTQFAKRTGTCLFLVG
                     HVTKEGALAGPRVLEHMVDTVLYFEGEQDSRFRLLRAVKNRFGAANELGIFAMTETGL
                     KTVSNPSAIFLSRYEDLQPGSVVMVAWEGTRPLLVEVQALVDESHSSNPRRIAVGLDQ
                     QRLAMLLAVLNRHGGIASYDQDVFINVVGGMKITETAADLALLLACVSSLRGKALSQE
                     LIVFGEIGLSGEIRPVQRGLERIKEAVKHGFTQAIVPLANCPKQKVDGIEIVGVKHLE
                     DALATL"
     gene            2834. .3322
                     /gene="ospA"
                     /evidence=experimental
     CDS             2834. .3322
                     /gene="ospA"
                     /note="similar to the rickettsial 17 kDa antigen"
                     /codon_start=1
                     /transl_table=11
                     /product="17 kDa antigen"
                     /protein_id="AAG17000.1"
                     /db_xref="GI:10441346"
                     /translation="MNRGCLQGSSLIIISVFLVGCAQNFSRQEVGAATGAVVGGVAGQ
                     LFGKGSGRVAMAIGGAVLGGLIGSKIGQSMDQQDKIKLNQSLEKVKAGQVTRWRNPDT
                     GNSYSVEPVRTYQRYNKQERRQQYCREFQQKAMIAGQKQEIYGTACRQPDGRWQVIST
                     EK"
     repeat_region   3560. .3846
     gene            complement(3842. .4561)
```

```
                     /gene="tnpA"
     CDS             complement(3842. .4561)
                     /gene="tnpA"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAG16999.1"
                     /db_xref="GI:10441345"
                     /translation="MDLTLISLFCVIDDFCQELLPQWNAILLEDTNKKRNKPSQMSTS
                     EIMTIMIYFHKSNYRNFKMYYLHVIKGSMVKYFPNSVSYNRFVELMPSILLPLCFFIA
                     AQGKTATGIYFVDSTILRVCHEKRASQNRAFKGLAKKSKSTMGWYYGFKLHIIVNDMG
                     ELMAFKMSKATTDDRVVLPKMAENLTGKIIGDKGYISQKLFDQLYEKGLQLITKIRKN
                     MKNKLVLMIDKILLKKRGIIA"
     repeat_region   4562. .4849
BASE COUNT     1362 a    778 c   1265 g   1578 t
ORIGIN

  Query Match             61.1%;  Score 295;  DB 1;  Length 4983;
  Best Local Similarity   76.0%;  Pred. No. 8.5e-62;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64
           | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||
Db    2841 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 2900

Qy      65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124
           ||||||| || || || ||||| || || || || || || ||||| ||||||||||| |
Db    2901 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 2960

Qy     125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184
           |||||||||| ||||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db    2961 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 3020

Qy     185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244
           | ||  | ||||| ||||||||||||||    ||||| ||||||||||||| || || |
Db    3021 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 3080

Qy     245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304
           |||||  | ||||||| || ||||| || || || || ||||||||||||| || || |
Db    3081 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 3140

Qy     305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364
           | || || ||   ||| || || || || || ||||||||||||| || || || || |
Db    3141 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 3200

Qy     365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424
           | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db    3201 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 3260

Qy     425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483
           | |||||||| || |||| || |||||||| || ||||| || |||   || ||||||
Db    3261 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 3319


RESULT 3
RIR17KGCA
LOCUS       RIR17KGCA                537 bp    DNA     linear   BCT 04-FEB-1999
DEFINITION  Rickettsia japonica 17K genus-common antigen gene, complete cds.
ACCESSION   D16515
VERSION     D16515.1  GI:695412
KEYWORDS    17K genus-common antigen; 17k dalton protein.
SOURCE      Rickettsia japonica (sub_species:YH) DNA.
  ORGANISM  Rickettsia japonica
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Furuya,Y., Katayama,T., Yoshida,Y. and Kaiho,I.
  TITLE     Specific amplification of Rickettsia japonica DNA from clinical
            specimens by PCR
  JOURNAL   J. Clin. Microbiol. 33 (2), 487-489 (1995)
  MEDLINE   95229950
REFERENCE   2  (bases 1 to 537)
  AUTHORS   Yoshida,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1993) Yoshiya Yoshida, Kanagawa Prefectural
            Pabulic Health Laboratoty; 52-2 Nakao-cho, Asahi-ku, Yokohama,
            Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Rickettsia japonica"
                     /sub_species="YH"
                     /db_xref="taxon:35790"
                     /note="propagated in Vero cells"
     CDS             58. .537
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA03965.1"
                     /db_xref="GI:1435056"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGTGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGQWQVVN"
BASE COUNT      176 a     92 c    129 g    140 t
ORIGIN

  Query Match              9.6%;  Score 46.6;  DB 1;  Length 537;
  Best Local Similarity   47.2%;  Pred. No. 0.62;
  Matches  176;  Conservative    0;  Mismatches  194;  Indels    3;  Gaps    1;

Qy     100 GGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCG 159
           || |  | ||  ||||    |     ||     |  |||||||| ||| | || |   |   |
Db     163 GGCGGTGCTGGTGGCGCATTACTTGGTTCTCAATTCGGTAAGGGCACAGGACAGCTTGTG 222

Qy     160 ATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATG 219
           |     |  |||||  | || || |   | ||||     ||||||||    | |||
Db     223 GGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTATG 282

Qy     220 GACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTT 279
           ||  |||||||||| ||         | |   |     | | ||    |   |     | |
Db     283 GATGAGCAGGATAGAAGACTTGCAGAGCTTACCTCACAGAGAGCTTTA---GAAACAGCT 339

Qy     280 ACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTAC 339
           || || |  ||||    ||     ||||     |    |   |    | | |  |
Db     340 CCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAACGGCAATTACGGTTACGTAACA 399

Qy     340 CAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCC 399
           |    | | |   |     | ||  |   || || ||||| || |     || | ||
Db     400 CCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGCCGTGAGTACACTCAAACAGTT 459

Qy     400 ATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGC 459
           | || |  ||  |  |||| ||   ||||| |  |||||||  || || || || |
Db     460 GTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCGTGCCGCCAACCTGACGGACAA 519

Qy     460 TGGCAGGTGATTA 472
           ||||| ||  | |
Db     520 TGGCAAGTTGTGA 532


RESULT 4
RPXX04
LOCUS       RPXX04                237523 bp    DNA     linear   BCT 11-NOV-1998
DEFINITION  Rickettsia prowazekii strain Madrid E, complete genome; segment
            4/4.
ACCESSION   AJ235273 AJ235269
VERSION     AJ235273.1  GI:3861237
KEYWORDS    complete genome.
SOURCE      Rickettsia prowazekii.
  ORGANISM  Rickettsia prowazekii
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; typhus group.
REFERENCE   1  (bases 1 to 237523)
  AUTHORS   Andersson,S.G., Zomorodipour,A., Andersson,J.O.,
            Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K.,
            Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
  TITLE     The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
```

```
  JOURNAL   Nature 396 (6707), 133-140 (1998)
  MEDLINE   99039499
   PUBMED   9823893
REFERENCE   2  (bases 1 to 237523)
  AUTHORS   Andersson,S.G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1998) S.G.E. Andersson,
            Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University
            of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .237523
                     /organism="Rickettsia prowazekii"
                     /strain="Madrid E"
                     /db_xref="taxon:782"
     gene            164. .1123
                     /gene="RP701"
     CDS             164. .1123
                     /gene="RP701"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15137.1"
                     /db_xref="GI:3861238"
                     /translation="MAEPIIFTLLFGNKDRILDRINNQKALNKLHIIPLQVNKESHIS
                     DKERIFSEILEKSYKECKIPIFNIQISNNNVKPIFRAQDLINAQNLNIKTTITFDKYN
                     ALPQNSELKSYWKQIIEKVDHVFFTNEADQNLSIADGIVPKDKTTTITDIRLVTSVFN
                     NLIYNHETNHGLSSTIYKKVDKLIKTAKNQDSKEIIETWPISVDEATNLITAKFGITS
                     EDQIYGLKLEIHEILKDPNNAAENFKKYVSQISRQFKKDLDKDQKNHIDSNFNTQKVI
                     NDSNKDIQVEHITSYVQDKEKQSKPQGFLTRIFNYFKNIVFKQ"
     gene            2074. .3348
                     /gene="RP702"
     CDS             2074. .3348
                     /gene="RP702"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15138.1"
                     /db_xref="GI:3861239"
                     /translation="MTTILVIVIIIMISLSALFAATETAITASSPGKIHKLKIAGNKR
                     AKTVLEVLKKKEKVIGTLLIGNSLINTICTTIATTLFISLLGNNGTIVASGVMAFIII
                     VFAEVVPKAIAVAKPEQLTLKMSSTIVIFLKLFKPINIALDYITKIFCFIFRINLNPQ
                     ISGTEEVRGVIEHYHQEGGVYKSERNMLGGILDIRNMTVSEIMTHRSNIIALNIDLPY
                     ELIIKTLLSGAHTRIPLWQDNRDNIIGILNLKDLLKALYENNNDDKKVDINTLLTPPW
                     FIPDNALVVDQLHAFRERNNHFACVVDEYGTLLGIITLEDVIEEIVGPITDEHDRLNN
                     EIIQESNTEFIIKGTTTIRDINRELDWNLSDEDANTIAGLIIHKIARIPNQGEVIKIF
                     NFKIIILKKIANKIDSVKITVLPMTEETISSE"
     gene            3341. .5353
                     /gene="RP703"
     CDS             3341. .5353
                     /gene="RP703"
                     /codon_start=1
                     /transl_table=11
                     /product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
                     (ccmF)"
                     /protein_id="CAA15139.1"
                     /db_xref="GI:3861240"
                     /translation="MSKIGNFLLLTTMCLSGILTLLIIPWLDYLIKKKHRNNICLKFF
                     PWIPCTSQVMTQAITFYASSLCSIFAFCTLVYAFITSDFSVQNVFLHSSTLKPLIYKI
                     SGSWSSHEGSMLLWFCLLQIVSCCYIFCLEDKMLKFSSIIILSAIQLLFSSFIYFTSN
                     PFHMFSFVPKEGLGLNPMLQDIALSIHPPLLYLGYVSYAVPFTIVCAIFLIAVISRLD
                     NNLPGYFTGYKITQLSDRITSIRLLKIFAGCGILFTTIGISLGSWWAYRELNWGGLWF
                     FDPVENISLFPWLSGIMLHHSIIVTTKTNRMQHWTIILSIITFLLVIFSTFLVRSGFI
                     NSIHAFAFSSERWIYLLVILCIIGIGGLYVIIRLNNCSIEKTIYNTFNWNTHSNCNMT
                     IKEKAILIGNIFFLFSLIILICATLYPTIYALFNDKPIIINEEFFIKNFIIFVIPILC
                     TIGLFTTNSSIKKHIIMLMLSLIITYFILLNIKFSVISILTIMTSIFLMLHNVHYLLV
                     KTNYFRNSLKASSTSMILGHFGVALIAFSITMNSLLQSEMDFIGEVGTSKTFNAFKVT
                     LQNIKFAQGKNYYRQIAEFWLEDHKNNITILKPENRFYIIEQKLSQESDIHSYLLYDL
                     YAVLSNIDGKIIHAKIYYKPMMSFIWIGVILTTAGFVIALIRKNSY"
     gene            complement(7264. .12195)
                     /gene="RP704"
     CDS             complement(7264. .12195)
                     /gene="RP704"
                     /codon_start=1
                     /transl_table=11
                     /product="CELL SURFACE ANTIGEN (sca5)"
                     /protein_id="CAA15140.1"
                     /db_xref="GI:3861241"
                     /translation="MAQKPNFLKKIISAGLVTASTATIVAGFSGVAMGAAMQYNRTTN
                     AAATTFDGIGFDQAAGANIPVAPNSVITANANNPITFNTPNGHLNSLFLDTANDLAVT
                     INEDTTLGFITNIAQQAKFFNFTVAAGKILNITGQGITVQEASNTINAQNALTKVHGG
                     AAINANDLSGLGSITFAAAPSVLEFNLINPTTQEAPLTLGANSKIVNGGNGTLNITNG
                     FIQVSDNTFAGIKTINIDDCQGLMFNSTPDAANTLNLQVGGNTINFNGIDGTGKLVLV
                     SKNGAATEFNVTGTLGGNLKGIIELNTAAVAGKLISQGGAANAVIGTDNGAGRAAGFI
                     VSVDNGNAATISGQVYAKNMVIQSANAGGQVTFEHIVDVGLGGTTNFKTADSKVIITE
                     NSNFGSTNFGNLDTQIVVPDTKILKGNFIGDVKNNGNTAGVITFNANGALVSASTDPN
                     IAVTNINAIEAEGAGVVELSGIHIAELRLGNGGSIFKLADGTVINGPVNQNALMNNNA
                     LAAGSIQLDGSAIITGDIGNGGVNAALQHITLANDASKILALDGANIIGANVGGAIHF
                     QANGGTIKLTNTQNNIVVNFDLDITTDKTGVVDASSLTNNQTLTINGSIGTVVANTKT
                     LAQLNIGSSKTILNAGDVAINELVIENNGSVQLNHNTYLITKTINAANQGQIIVAADP
                     LNTNTTLADGTNLGSAENPLSTIHFATKAANADSILNVGKGVNLYANNITTNDANVGS
                     LHFRSGGTSIVSGTVGGQQGHKLNNLILDNGTTVKFLGDTTFNGGTKIEGKSILQISN
                     NYTTDHVESADNTGTLEFVNTDPITVTLNKQGAYFGVLKQVIISGPGNIVFNEIGNVC
                     IVHGIAANSISFENASLGTSLFLPSGTPLDVLTIKSTVGNGTVDNFNAPIVVVSGIDS
                     MINNGQIIGDKKNIIALSLGSDNSITVNANTLYSGIRTTKNNQGTVTLSGGMPNNPGT
                     IYGLGLENGSPKLKQVTFTTDYNNLGSIIANNVTINDYVTLTTGGIAGTDFDAKITLG
                     SVNGNANVRFVDSTFSDPRSMIVATQANKGTVTYLGNALVSNIGSLDTPVASVRFTGN
                     DSGAGLQGNIYSQNIDFGTYNLTILNSNVILGGGTTAINGEIDLLTNNLIFANGTSTW
                     GDNTSISTTLNVSSGNIGQVVIAEDAQVNATTTGTTTIKIQDNANANFSGTQAYTLIQ
                     GGARFNGTLGAPNFAVTGSNIFVKYELIRDSNQDYVLTRTNDVLNVVTTAVGNSAIAN
                     APGVSQNISRCLESTNTAAYNNMLLAKDPSDVATFVGAIATDTSAAVTTVNLNDTQKT
                     QDLLSNRLGTLRYLSNAETSDVAGSATGAVSSGDEAEVSYGVWAKPFYNIAEQDKKGG
                     IAGYKAKTTGVVVGLDTLASDNLMIGAAIGITKTDIKHQDYKKGDKTDINGLSFSLYG
                     SQQLVKNFFAQGNAIFTLNKVKSKSQRYFFESNGKMSKQIAAGNYDNMTFGGNLIFGY
                     DYNAMPNVLVTPMAGLSYLKSSNENYKETGTTVANKRINSKFSDRVDLIVGAKVAGST
                     VNITDIVIYPEIHSFVVHKVNGKLSNSQSMLDGQTAPFISQPDRTAKTSYNIGLSANI
                     KSDAKMEYGIGYDFNSASKYTAHQGTLKVRVNF"
     gene            complement(12953. .13216)
                     /gene="RP705"
     misc_feature    complement(12953. .13216)
                     /gene="RP705"
                     /note="Pseudogene that lacks ATG or GTG initiation codon.
                     Homology to guanosine-3,5-bis(diphosphate)
                     3-pyrophosphohydrolase."
     gene            complement(13563. .14507)
                     /gene="RP706"
     CDS             complement(13563. .14507)
                     /gene="RP706"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15141.1"
                     /db_xref="GI:3861242"
                     /translation="MKIRQTAKPVIIGISGPELTDAERELFEEHNPLGIILFRRNIRK
                     NEKGEQDKAALIKLIADIKEVLGDNTIIAIDQEGGMVKRLIAPTFYDAPAAQTFTELQ
                     TCKYNYSTIAKELREVGINLDFAPVADLIHDGADKIISDRSFGKEPEIVVPLFLSAIA
                     GLQEEKVTACIKHIPGHGRATVDSHIELPIIDTSLKTLEDTDFKVFKELAKYDYIKLA
                     MTAHVIYTALDPDNPATLSQIVIDYIKSNIGFKGLIISDTIEMQALSGSMADITKGAL
                     DAGVDIVLECTGEFNKMSEVLGSVAENSIDKFSDLFIS"
     gene            complement(14552. .15025)
                     /gene="RP707"
     CDS             complement(14552. .15025)
                     /gene="RP707"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15142.1"
                     /db_xref="GI:3861243"
                     /translation="MINNKLTKKYYVASEVIRILNIALHKLRYLETKIPGLSHYKINN
                     RKYYTAHDIDLLQKFLFVSRREEYLEVRGYHRSDKDYLANFSNIHCASDTSRQNEFQG
                     KNSECILNLHVQQRQDSQSAVVLSLMDNTVSSRIDILLTNFHNLSIQIKKILLIS"
     gene            complement(15220. .15507)
                     /gene="RP708"
     CDS             complement(15220. .15507)
                     /gene="RP708"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (himA)"
                     /protein_id="CAA15143.1"
                     /db_xref="GI:3861244"
                     /translation="MTITKNKISLMLSSKLGFSNNLCEAIVNTVFSNILEIAKVQKLT
                     LKNFGSFEVKQKTPRPGINFHTKSPVMIASKKNLRFTPSEKLKALINKSML"
     gene            18405. .18701
                     /gene="RP709"
     CDS             18405. .18701
                     /gene="RP709"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15144.1"
                     /db_xref="GI:3861245"
                     /translation="MAVGRQYIEDHTEQLHKLCDNLVKYIELETKAITHEKAAEDTEV
                     VIEVSLDLDNSSDDASIINKKAITVTSDDIANSNEDAARNTNKSHGDVLRCYYQ"
     gene            complement(19685. .20789)
                     /gene="RP710"
     misc_feature    complement(19685. .20789)
                     /gene="RP710"
                     /note="Pseudogene with two authentic frameshifts and four
                     authentic in-frame termination codons. Homology to
                     transposase."
     gene            complement(21112. .21608)
                     /gene="RP711"
     misc_feature    complement(21112. .21608)
                     /gene="RP711"
                     /note="Pseudogene with one authentic frameshift. Homology
                     to resolvase."
     gene            complement(22602. .23066)
                     /gene="RP712"
     CDS             complement(22602. .23066)
                     /gene="RP712"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15145.1"
                     /db_xref="GI:3861246"

  Query Match             9.3%;  Score 45;  DB 1;  Length 237523;
  Best Local Similarity   46.9%;  Pred. No. 1.6;
  Matches  175;  Conservative    0;  Mismatches  195;  Indels    3;  Gaps    1;

Qy    100 GGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCG 159
          ||||  |  |  || |   |    ||    |  |||||| ||||    || |   |
Db 172645 GGTGGCGCAGGAGGTGCATTACTTGGTTCTCAATTCGGTCAAGGTAAAGGACAACTTGTC 172704

Qy    160 ATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATG 219
            |      |  || ||  | || || |    |  ||||     |||||||    || |||
Db 172705 GGAGTAGGTGTAGGCGCATTACTTGGGGCAGTTCTTGGTGGACAAATCGGCGCAAGTATG 172764

Qy    220 GACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTT 279
          ||  |||||||||| ||     ||  | |   |     ||| ||    |   |     |
Db 172765 GATGAGCAGGATAGAAGACTTCTAGAACTAACATCACAAAGAGCTTTA---GAATCTGCA 172821

Qy    280 ACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTAC 339
           || |  |  ||||    ||     | ||     |    |  | |  | |  |  |
Db 172822 CCTAGCGGTAGTAACATAGAATGGCGCAATCCAGATAACGGCAATCATGGTTACGTCACA 172881

Qy    340 CAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCC 399
          |    | | |    |   | |  |        || || ||||| ||||     || | ||
Db 172882 CCTAATAAAACTTATAGGAACAGTGCAGGTCAATATTGCCGTGAATACACTCAAACAGTT 172941

Qy    400 ATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGC 459
          || || |  ||  |  |||| |||  || || |  || |||| ||| || ||||| |
Db 172942 ATAATAGGCGGAAAACAACAAAAAACATATGGTAATGCATGCCGTCAACCTGATGGGCAA 173001

Qy    460 TGGCAGGTGATTA 472
          ||||| ||  | |
Db 173002 TGGCAAGTTGTGA 173014


RESULT 5

AL646071
LOCUS       AL646071              203050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 15/19.
ACCESSION   AL646071 AL646052
VERSION     AL646071.1  GI:17429627
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 203050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .203050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            complement(180. .1025)
                     /gene="RSc2606"
                     /note="RS00854"
     CDS             complement(180. .1025)
                     /gene="RSc2606"
                     /function="elements of external origin; plasmid-related
                     functions non-bacterial functions"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL REPLICATION INITIATOR AND
                     TRANSCRIPTION REPRESSOR PROTEIN"
                     /protein_id="CAD16313.1"
                     /db_xref="GI:17429628"
                     /translation="MSSPPGQALPQREQLDLFRALPGDMAPRDSQDLMAFPFFSLAKS
                     RRTAPIDFRAGGITIRVEGTAEHGIATIWDADVLIWAASQIVEARDAGIRPSRWIRAT
                     PYEILRFIGRSTSLNDYQRLKAALDRLQSTSIATSIRETTGRRLHRFSWINEWRELAD
                     ASGKPLGIELILPDWFYAGVLDAALVLTIDPAYFRLKGGIERWLYRLVRKHGGRQEHG
                     WQFDFKHLHRKSGSAARFSDFAYDLRALVARQSLPGYVLGIERMPDGGQELLTFRPVL
                     HTARG"
     gene            complement(1052. .1327)
                     /gene="RSc2607"
                     /note="RS00855"
     CDS             complement(1052. .1327)
                     /gene="RSc2607"
                     /function="miscellaneous; hypothetical/partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
```

predicted by Codon_usage
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD16314.1"
/db_xref="GI:17429629"
/translation="MRPAPLRPAAAAPAQPQRYLTNDEAAEYLRLSPRTLEKQRVVGG
GPKFRKFGRRVMYAVSDLDAWADQRSYEATSDPEYAERHAGDYRDGR"

gene complement(1442. .2212)
/gene="RSc2608"
/note="RS00857"

CDS complement(1442. .2212)
/gene="RSc2608"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAD16315.1"
/db_xref="GI:17429630"
/translation="MTDRSAEHWYPTAAYLYVLHLDGPALAWEYLRRNPDYRRDWLRR
RRRPDAAQVWGLRLLEDPALDARDAHPAWFPDHDAVVQLYPDADPPPEAHAFEFWRVP
GRKQLIHDGKRLVLVSHWPGCCLRLALAPGLEDGMAYLYATRACATPCARYRTLTAEL
DALATATVAAPTAAARSRPTLAAVLELHTLQALDATLAGASLREVAVGLFSADAVAAD
WHKDSALRARVRRLVRRGEALMRGGYRRLAQLPLLPLQ"

misc_feature 2213. .8108
/note="ACUR or RS00937
predicted by FrameD"
/evidence=not_experimental

gene complement(2557. .3150)
/gene="RSc2609"
/note="RS00858"

CDS complement(2557. .3150)
/gene="RSc2609"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE LIPOPROTEIN"
/protein_id="CAD16316.1"
/db_xref="GI:17429631"
/translation="MTERPLSCFYGNPHLRIPTEPGNRIRVFPASDNSASVLPFFCGC
DGLHIRASTRMHRYGSTVLCFGTRAHPRIVGFVRRLPMTLPLVTADERARLLAHGAAL
AAGQRLDPLPVVRLFTPDAHVTWLLVSLDPADGDTAYGLIDLGIGLPSLGTVKLSDLA
GIVGPRKLPVRRDRYFQAAHLLSEYVRLAQENGAITD"

gene 3223. .3513
/gene="RSc2610"
/note="RS00859"

CDS 3223. .3513
/gene="RSc2610"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD16317.1"
/db_xref="GI:17429632"
/translation="MAAKYSLAKALKTVRKARGLSQEAFSDVSSRTYMSTLERDLKSP
TLHKLTELCEVMDVHPLTLLTLAFAGDSTRKADELLAQVRQELEAVLNSDGD"

gene 3522. .4085
/gene="RSc2611"
/note="RS00860"

CDS 3522. .4085
/gene="RSc2611"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAD16318.1"
/db_xref="GI:17429633"
/translation="MAYRNGNYTAFYVAEPFHPSSLGANATKDFQYYNTLRMWKGADA
TFPFVDSHDKTYSVRDGSDWEYTLKPRLRERLRSSKNIILILSSITTNSRALREEIDY
GINDQGLPVIVIYPEYATKESLLVDGSLKQAVKNLWGNLPVFRDSMSKIPTLHVPMNK
ELIKKSLQDGEFMVGSMKAPDIYRYNP"

gene 4107. .4955
/gene="RSc2612"
/note="RS00912"

CDS 4107. .4955
/gene="RSc2612"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD16319.1"
/db_xref="GI:17429634"
/translation="MIHVMAKVNFFDKRVIKKFLEITAVVSGTLSFIVIFVDVPAEWK
LKAGLAFLAILALTYVVIWLWSNNLNSIDINVEGSDVAIKTGDIFQQPGLKAIAFNEY
FDTQVDNKIIGETSLNGVFIQKHLGVPISELDRYIEEYAFDESEVLEENEARKYGKKK
RYQIGTICLYKDYLLTAFSKFDENNKALLTMPEYLEFLINFWDKVNNVYAQQSVSTTI
FGSGITRIKGHKNISDEDLLKIMLWTFRISEMRFKYPAKLTIVIHKDKIDQINLLDIK
SARNGV"

gene complement(5305. .5619)
/gene="RSc2613"
/note="RS00913"

CDS complement(5305. .5619)
/gene="RSc2613"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"

Query Match 8.9%; Score 42.8; DB 1; Length 203050;
Best Local Similarity 56.3%; Pred. No. 5.5;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy     327 GGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATT 386
           | || ||  || |    |   | |  |||||  ||          ||||||     |
Db  105900 GCTTGGCGTCTGCGCCTGCGCGATCGCGCAGGACATCCCGGTCGGCACCTGCCGGCCGAT 105959

Qy     387 TCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCA 446
             |  | |||||  |||||  ||   ||  ||| |  ||  |||||| |||||||  ||
Db  105960 GGTGGGGCAAGCCGAGATCGACGGCGTGATGCAGCAGGTCAGCGGCACGGCGTGCCTGCA 106019

Qy     447 GCCGGATGGCCGCTGGCAGGTG 468
           |||||| |||  ||||||| ||
Db  106020 GCCGGACGGCACCTGGCAGATG 106041
```

```
RESULT 6
RSU76907
LOCUS       RSU76907                448 bp    DNA     linear   BCT 18-FEB-1998
DEFINITION  Rickettsia sp. 17kDa common-antigen gene, partial cds.
ACCESSION   U76907
VERSION     U76907.1  GI:2894794
KEYWORDS    .
SOURCE      Rickettsia sp.
  ORGANISM  Rickettsia sp.
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Davis,M.J., Ying,Z., Brunner,B.R., Pantoja,A. and Ferwerda,F.H.
  TITLE     Rickettsial relative associated with papaya bunchy top disease
  JOURNAL   Curr. Microbiol. 36 (2), 80-84 (1998)
  MEDLINE   98087556
   PUBMED   9425244
REFERENCE   2  (bases 1 to 448)
  AUTHORS   Ying,Z. and Davis,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1996) Tropical Research and Education Center,
            University of Florida, 18905 SW 280 Street, Homestead, FL 33031,
            USA
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Rickettsia sp."
                     /db_xref="taxon:789"
                     /note="Puerto Rico"
     CDS             <1. .>448
                     /note="putative"
                     /codon_start=3
                     /transl_table=11
                     /product="17 kDa common-antigen"
                     /protein_id="AAC02809.1"
                     /db_xref="GI:2894795"
                     /translation="ALAASMLEACQGPGGMNKQGSGTLIGGTLGGLVGSQFGGGTGRL
                     AAVGAGALLGAILGNQIGAGMDEQDRKLAELTSQRALEAAPSGSSVQWRNPDNGNYGT
                     VTPSKAYKNNTGQYCREYTQTVVVGGKQQKAYGTACRQPDGQWQVV"
BASE COUNT      135 a     81 c    120 g    112 t
ORIGIN

  Query Match             8.6%;  Score 41.6;  DB 1;  Length 448;
  Best Local Similarity   47.7%;  Pred. No. 10;
  Matches  184;  Conservative    0;  Mismatches  199;  Indels    3;  Gaps    2;

Qy      87 TGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTC 146
           |||  ||   |  || |    | | || ||  |||  ||    |  || |||   ||  |
Db      65 TGGTACGCTAATTGGCGGAACTTTAGGAGGACTTGTTGGTTCACAATTTGGTGGTGGTAC 124

Qy     147 TGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAAT 206
           || ||  |  |   ||     |  |||||| | || || |   |  |||||    ||||
Db     125 GGGACGACTTGCTGCGGTAGGAGCAGGTGCGTTACTTGGTGCAATCCTTGGCAACCAAAT 184

Qy     207 CGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAA 266
           |||    | |||||  | || |||| ||          | |   ||    |||  ||
Db     185 TGGTGCAGGTATGGATGAACAAGATAGAAAACTTGCAGAGCTTACTTCTCAAAGAGCTTT 244

Qy     267 AGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAACC 326
           ||   ||       |  || | ||    | |      |       | | || |       |
Db     245 AGA-AGCTGCACCAAGCGGTAGCAGTGTACAATGGCGTAATCCTGATAATGGTAATTATG 303

Qy     327 GGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATT 386
           |    |   |  | |||    |    || |  ||  |||    || || || || ||||
Db     304 GTACTGTGACACCAAGCAAAGCTTATAAAAATAATACCG--GTCAATATTGTCGTGAATA 361

Qy     387 TCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCA 446
           || | ||   |  | |  ||  |  ||||  ||   || ||||| || |||| |||
Db     362 CACCCAAACAGTTGTAGTAGGTGGAAAACAACAAAAAGCTTATGGCACTGCCTGCCGTCA 421

Qy     447 GCCGGATGGCCGCTGGCAGGTGATTA 472
           || ||||| |  ||||| ||  | |
Db     422 ACCTGATGGACAATGGCAAGTTGTGA 447
```

```
RESULT 7
AL646061
LOCUS       AL646061              198050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 5/19.
ACCESSION   AL646061 AL646052
VERSION     AL646061.1  GI:17427781
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 198050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .198050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            151. .642
                     /gene="RSc0770"
                     /note="RS05084"
     CDS             151. .642
                     /gene="RSc0770"
                     /EC_number="1.-.-.-"
                     /function="small molecule metabolism"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (SMALL
                     SUBUNIT) OXIDOREDUCTASE PROTEIN"
                     /protein_id="CAD14472.1"
                     /db_xref="GI:17427782"
                     /translation="MHTRPPSSDGDDPAGLTRRQWLQSALAATAASLVGSVTLYAVAQ
                     APGEPLDAFMGLSQALTARSALDRAVGARLLAALQQASAGFAAQLRPLAQSLAAGTLS
                     GLQQETALRILQAWYTGVVDGTVVTYEQALMYGVVSDILVIRTYCPNRPGFWAEPPVE
                     KQS"
     gene            642. .2261
                     /gene="RSc0771"
                     /note="RS05083"
     CDS             642. .2261
                     /gene="RSc0771"
                     /EC_number="1.-.-.-"
                     /function="small molecule metabolism"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
```

predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (LARGE
SUBUNIT) OXIDOREDUCTASE PROTEIN"
/protein_id="CAD14473.1"
/db_xref="GI:17427783"
/translation="MAETRQAEQADIVVVGSGVAGALVAYELARAGKSVLMLEAGPRL
PRWEIVERFRNQADKMDFMAPYPSTPWAPHPEYGPDNNYLILKGEHKFDSQYIRAVGG
TTWHWAASTWRFLPNDFKLRSVYGIARDWPIQYDDLERYYGRAEAALGVWGPNDEELG
SPRSQPYPMAPLPLSFNERTVKDALNAHDPALRVVTEPVARNSRPYDGRPTCCGNNNC
MPICPIGAMYNGIVHVEKAEQAGARLIENAVVYKLEVGAGKRIVAAHYKDPKGVPHRV
EGKWFVLAANGIETPKLMLMSTGADFPKGVGNRSDMVGRNLMDHPGTGVSFYADRKLW
PGRGPQEMTSLIGFRDGPFRATQAGKKLHLSNISRIEQETTRIFKAGKLLKPAELDAR
IRDQAARYVQFDSFHEILPLPENRIVPSATGTDALGIPRPEITYRIDDYVKRSAVHTR
EVYATAARVLGGTDVQFHDDFAPNNHITGATIMGSDPKDSVVDKDCRTFDHPNLFISS
SATMPTVGTVNVTLTIAALALRIADQLKKEA"
gene 2263. .3552
/gene="RSc0772"
/note="RS05082"
CDS 2263. .3552
/gene="RSc0772"
/EC_number="1.-.-.-"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE DEHYDROGENASE
(CYTOCHROME C SUBUNIT) SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14474.1"
/db_xref="GI:17427784"
/translation="MRRAKAASFLLLSVLGAAAAARDADAPADAAQIKRGEYLAIAGD
CVACHTAPGGKPFAGGLPLQIPMLGTIYSSNITPDPKTGIGKWRYEDFERAVRQGVSD
DGENLYPAMPYPSYAKINDADMRDLYAYFRYGVAAVQYAPPPGTIRWPLNMRWPLKLW
NLAFLKRAPYLPKGDMTPTWNRGAYLAQGLAHCGTCHTPRGFAMQEKAMDERGKGYLA
GSTLAGWSAYNITSDPVSGIGGWKPEQIVQYLRTGSVPGIGQAAGPMGEAVQHSFSRM
TERDVRAIAEYLRTVPAVSTGAQRARHGWGTPATDVIALRGKPIETTIDAARLYLGNC
ATCHQADGRGTPDGYYPPLLHNSAVGARDTRNLVQVMLHGIERRAGDRHIGMPAFGRQ
LSDAQLAALANYVTRQFGDPATPALTADEIAKRRMPQ"
gene complement(3559. .5781)
/gene="RSc0773"
/note="RS05081"
CDS complement(3559. .5781)
/gene="RSc0773"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD14475.1"
/db_xref="GI:17427785"
/translation="MRTSARVILGLGITPIVIVAAAIVFVLTFDWNRAKPYLNDHVSQ
AVGRPFAINGDLALTWKKPEGESGWRAYVPWPRLIASDIAIGNPDWARQPHVATVQRL
TFVLEALPLLAHRIVIPSVTLDTPMIALERDSQNRNNWTFHFGAQDSQPSDWKLELRE
IAFSKGSVQLDDEVKRIHLAATLDTVDNQALYTAASGTAIRGSEPAAPSASGAAASTA
PAAPDPQQPYGIAWTVKGTYNNAAIRGSGKAGGVLRLQDARRPYPLQADVTVGKTRID
LAGTLTNPSSLTALDLRLHLSGASMAHLYPLTGVVLPDTPPFDTRGRLIGELRRQGSS
WRYEQFTGRVGGSDLGGTLTFAMRPEPGQRPQLTGELVSHQLLFADLAPIIGADSNAS
KQRRDAPVRQPADKVLPVEPFRTDRWNAIDADVKFTGERIVRTADLPIDHLVTHIKLQ
DAVLTLDPLNFGVAGGTLTSNLRLDGQSAPMQASAALAARHFKIRQLFPNIESMHASI
GEINGDARLSATGNSVAALLGASNGELKILVEQGTVSKFILEAMGLNVGNVILTKLFG
DKQVSINCAAGDFAVSDGLAQARTFVVDTQDAVIDVTGATSFKHESLDFTIHPDSKGL
RVFSLRTPLYVKGTYKHPDVSVNPAVVALRAGGAVALAFAAPVAAVLPVLELQPAPDS
PCGKLLADVRQRPTAPPPGKVYRSGQTGRAADGSPKDPAAAPARQPALPLRDPTTAGG
"
gene 5883. .6785
/gene="RSc0774"
/note="RS05080"
CDS 5883. .6785
/gene="RSc0774"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD14476.1"
/db_xref="GI:17427786"
/translation="MSHVAISAPAHPPALDDEYARRFGGVARLYGPQALERFAAAHIC
VIGVGGVGSWVAEALARCGIGRLTLIDLDHIAVSNTNRQIHALGDAYGKAKVDAMAER
IAEINPRAEISRIDDFVTVENVDALLGHPFDYVVDAIDAVKVKTAIVAFCQRTGTRVV
TCGAAGGQLDPTRIRVADLARTIQDPLLAKVRGNLRRQHGFPRDPKARFGIDAVYSDE
PLRYPEPEQQACAVEVPAESSHAGIDDLAAAGAIEVVNAPPAPQPSRGPQGLACAGFG
SSVCVTAVFGMVAASTALRAIAAV"
gene complement(6890. .8257)
/gene="katGb"
/note="RSc0775; RS05079"
CDS complement(6890. .8257)
/gene="katGb"
/EC_number="1.11.1.6"
/function="cell processes; protection responses;
detoxification"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE PEROXIDASE/CATALASE (CATA)
OXIDOREDUCTASE PROTEIN"
/protein_id="CAD14477.1"
/db_xref="GI:17427787"
/translation="MGPEPEAAGIEEQGLGWSSRFGTGKGTDAITSGLEVTWTTTPTQ
WSNNFFQNLFGYEWELTKSPAGAHQWVAKDAAETIPDAHDPSAKHLPTMLTTDLSLRF
DPAYEKISRRFYEHPEQFADAFARAWFKLTHRDMGPRARYLGPEVPAEALIWQDPIPA
VDHKLIDTQDIAALKARILASGLSVSQLVSTAWASASTFRGSDMRGGANGARIRLAPQ
KDWEANQPAQLAKVLETLESIQGAFNGAQSGGKKVSLADLIVLAGCAGVEQAAKNAGH
AVEVPFTPGRMDAAQAQTDVESFAVLEPIADGFRNYQKGKYTLPAEALLVDKAQLLTL
TAPEMTVLVGGLRVLDTNIGQTRHGVFTQRPESLTNDFFVNLLDMGTEWKATDGRDVF
EGRDRATGALKWTGTRVDLVFGSHSQLRALAEVYGSADAQAKFVRDFVAAWDKVMNLD

Query Match 8.5%; Score 41; DB 1; Length 198050;
Best Local Similarity 54.2%; Pred. No. 15;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120
Db 135919 GCTCCGGCCAACAGCCCCGGTATTCTCGGCACCATCATCGGCGGCGTGGCGGGTGGCGTG 135978

Qy 121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
Db 135979 CTGGGGCACCAGATCGGTGGCGGTACCGGCAATACGGTGGCCACCATCGGCGGCGCGGCC 136038

Qy 181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAG 213
Db 136039 CTGGGTGCCCTGGCGGGCAATCAGGTGGAGCGG 136071

```
RESULT 8
AL646079
LOCUS       AL646079              213050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 4/11.
ACCESSION   AL646079 AL646053
VERSION     AL646079.1  GI:17430956
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 213050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .213050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
                     /plasmid="megaplasmid"
     gene            100. .1632
                     /gene="RSp0487"
                     /note="RS00365"
     CDS             100. .1632
                     /gene="RSp0487"
                     /function="macromolecule metabolism; macromolecule
                     degradation; degradation of proteins, peptides,
                     glycopeptides"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD17638.1"
                     /db_xref="GI:17430957"
                     /translation="MDAPFSGRGAVIRRAVGRLLLVLVVGLAVLAVCIAFNTWRHDSR
                     QLDVPPAAPLAVDGAAAAARLGEAIRARTISSFTDAQQNADQFRQLHAMLAARYPKAH
                     AAMQREPVGDFSLLYTWKGSDPSLKPILLMAHQDVVPVAPGTEGDWTEPPFAGVVKDG
                     MVWGRGAWDDKGNLIAQMEAAELLAASGFRPRRTIHFAFGADEEVGGERGAARIAALL
                     KSRGEQLAFVIDEGLLITEGVVPGLTKPAALIGVTEKGFLSVALKLSATPGHSSMPPA
                     PGESAIAMMSAALKHLDDQQLPAGIRGVTREMFETLAPEMRGLNRVALSNLWLFEPLV
                     RKQLQASPSTNAVLQTTTALTIVQAGNKDNVLPGRAEATVNFRLLPGDSASSVIAHVE
                     QAVRSAVPKGHFELAALPGVSEAAPVSPTQSASYQLIGRTVREVFPGTVVAPGLMVGA
                     TDSRHMIGISDHVFRFSPVRARPEDLARFHGTNERISEANLVELIRFYHRLIHQAADA
                     GA"
     gene            complement(1751. .3148)
                     /gene="czcS"
                     /note="RSp0488; RS00366"
     CDS             complement(1751. .3148)
                     /gene="czcS"
                     /EC_number="2.7.3.-"
                     /function="cell processes; protection responses"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
                     HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
                     TRANSCRIPTION REGULATOR PROTEIN"
                     /protein_id="CAD17639.1"
                     /db_xref="GI:17430958"
                     /translation="MMRWPASLAARLVLAFVLVAGVMLTGAGTYLYYALNAQLRATDD
                     AMLLGKIVQLRHCAKDTQSIDDIVRDARRYTEVVSGHEGLILRVIGPDGRTLVEANPE
                     GEPVAAPAAVPPDVLASESALHTWWSRTGIASRAIAAQARLGNHRDTVTVLVAQDGDS
                     RAALMASYRRHIFLAVIAGVAAVALLSLWLVRSTLKGLARIGRQAASVVPSQLDTRLT
                     LEGAPRELSDLLAALNAMLTRLQDGFARLSQFSADLAHDFRTPLANLIGQTEVTLAHP
                     RPAGEYRAVLESSLEEYARLSRMIEDMLFLARADHARVALSVAPLDARHEIQVLVDYF
                     EAVAADRDITIQVCGNGRVRADLTLLRRAINNLLDNALRHTPGGARIDVAVQVLPHST
                     AIAVRNPGPGIPAASLPLLFDRFYRGDPARANSERSAGLGLAIVQTIMDLHGGTAQAS
                     SPPGGLTEFRLIFPA"
     gene            complement(3145. .3819)
                     /gene="czcR"
                     /note="RSp0489; RS00367"
     CDS             complement(3145. .3819)
                     /gene="czcR"
                     /function="cell processes; protection responses"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
                     CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
                     /protein_id="CAD17640.1"
                     /db_xref="GI:17430959"
                     /translation="MRLLIVEDEPKAGDYLLKGLTESGFVADLARSGPDGLYQAMEHD
                     YDLIVLDVMLPGMDGWQVIRELRRRKTTPVLFLTARDELSDRLKGLELGADDYLVKPF
                     AFAELVARVHTILRRGPMRESEVLEIADLTIDTIKRRVTRAGRRIDLTAKEYALLHLL
                     ARRTGEVLSRSLISSQVWDVNFDSNTNVVDVAIRRLRAKVDDPFEHKLIHTLRGMGYV
                     LDTVAR"
     gene            4089. .4394
                     /gene="RSp0490"
                     /note="RS00368"
     CDS             4089. .4394
                     /gene="RSp0490"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
                     /protein_id="CAD17641.1"
                     /db_xref="GI:17430960"
                     /translation="MSVRRNCLLPLAAAAAFLACGAATAASVSEQALSPLQVHGARDP
                     YTDGARGTRDPYTDGGAQAEATLQASVGPRDPYSDGGMSRSSAPRDRDPYTDGAQAA"
     gene            4400. .5725
                     /gene="czcC"
                     /note="RSp0491; RS00369"
```

```
     CDS             4400. .5725
                     /gene="czcC"
                     /function="cell processes; protection responses"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
                     RESISTANCE PROTEIN"
                     /protein_id="CAD17642.1"
                     /db_xref="GI:17430961"
                     /translation="MVTHPCIRAAVALSCALLVRSGLAAAPAGEAAAASDTAAPAAQS
                     APASVPGAAHERMTLQQAIDRALATHPALAAAGYAVRASEGAVDQASRLRNPEAAFLT
                     EDVRRDRATMTAQLNIPLELGGKRAARTRAAELARDAAGQDAAVLAAELRADVTRAFF
                     ELSIAQERLRLAQASDALAAQAEAAAVRRVQAGKVSPVEQTKAAVARAGAAIELRDAQ
                     ADLLVAQRALSAFWGGAVAVPEADGSAETLPAVPPEAHRVLAAATAAEPVALASTPRA
                     LRARLEIEHRQALVNVERSKRVPDVTLSVGAKRDSGANANMAVVGIAVPLPLFDRNQG
                     SLLEAQRLADKAAEDYRALRLEQAAALSQDVARLEAARQAVGALRQDVLPGAGRAYDA
                     ARIGFEAGKFNFLDVLDAQRTLFQARAQYLAALSRAHQAAAGIDRILGR"
     gene            5752. .6966
                     /gene="czcB"
                     /note="RSp0492; RS00370"
     CDS             5752. .6966
                     /gene="czcB"
                     /function="cell processes; protection responses"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE COBALT-ZINC-CADMIUM RESISTANCE (CATION
                     EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"
                     /protein_id="CAD17643.1"
                     /db_xref="GI:17430962"
                     /translation="MTLPRQQRRAIALIVALGALAGAALWWGMGARQAAHAESEAEQP
                     AAHAEHAQAPKAEAAEAPHAETIAMTAEAIANAAIGVETAGAAQLRTEVVLPGEIRLN
                     DDRTAHVVPRVAGVAEQVAVELGQPVTRGQLLAVLSSPALSDQRSELQAAQQRLALAQ
                     EAHAREKRLWEQGIAAAQDYQQARTALQEARIAADNARQKLAAIGAAPAGTALNRFEL
                     RAPFDGVVVARHLSQGEAVQAEAAVFTIADLRTVWADFAVTAKDLEAVTTHATATVRA
                     TATGTAVQGKVSYVGALLGEQTRSAPARVTLDNPKRAWRPGMFVSVSVESGRVPVPVA
                     VRADAVQTVDEKPSVFVPVPAGFEVRPVKTGRSDGTWTEILDGLPAGTRYAAANSYVL
                     KAEAGKSADHGH"
     gene            6977. .10171
                     /gene="czcA"
                     /note="RSp0493; RS00371"
     CDS             6977. .10171
                     /gene="czcA"
                     /function="cell processes; transport of small molecules"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage

  Query Match              8.4%;  Score 40.8;  DB 1;  Length 213050;
  Best Local Similarity   58.1%;  Pred. No. 17;
  Matches   72;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy      85 GTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGC 144
           || |||||||     || |  || || |||||  | | ||| || | | |||||   ||
Db   20248 GTGGGCGCGGTGCTGGGCGGCGTGGTCGGCGGGCTGGTCGGACACCAGATCGGTGGCGGG 20307

Qy     145 TCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAA 204
             ||   |  |  |    |||||||||| || ||   |||||  |||   |||   |
Db   20308 CGCGGCAATACGGTGGCCACCATCGGCGGGGCTGTGGCGGGCGCGCTGGCCGGCAACACC 20367

Qy     205 ATCG 208
           ||||

Db  20368 ATCG 20371


RESULT 9
RIRANT17KA
LOCUS       RIRANT17KA               539 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.conorii 17K genus-common antigen gene, complete cds.
ACCESSION   M28480
VERSION     M28480.1  GI:152457
KEYWORDS    antigen.
SOURCE      R.conorii DNA.
  ORGANISM  Rickettsia conorii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Anderson,B.E. and Tzianabos,T.
  TITLE     Comparative sequence analysis of a genus-common rickettsial antigen
            gene
  JOURNAL   J. Bacteriol. 171 (9), 5199-5201 (1989)
  MEDLINE   89359171
   PUBMED   2768201
COMMENT     Draft entry and printed copy of sequence for [1] kindly submitted
            by B.Anderson, 27-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Rickettsia conorii"
                     /db_xref="taxon:781"
     -35_signal      1. .6
     -10_signal      21. .27
     mRNA            38. .>539
                     /product="17 kD mRNA"
     CDS             60. .539
                     /note="17 kD antigen precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26376.1"
                     /db_xref="GI:152458"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGKGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGQWQVVN"
     sig_peptide     60. .119
                     /note="17 kD antigen signal peptide"
     mat_peptide     120. .536
                     /product="17 kD antigen"
BASE COUNT      179 a     93 c    125 g    142 t
ORIGIN

  Query Match              8.3%;  Score 40.2;  DB 1;  Length 539;
  Best Local Similarity   46.1%;  Pred. No. 23;
  Matches  172;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy     100 GGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCG 159
           || |  | ||  ||||   |    ||    |  |||||||| |||   || |   |
Db     165 GGCGGTGCTGGCGGCGCATTACTTGGTTCTCAATTCGGTAAGGGCAAAGGACAGCTTGTT 224

Qy     160 ATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATG 219
             |     |  |||||  | || || |   |  ||||     ||||||||    | |||
Db     225 GGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTATG 284

Qy     220 GACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTT 279
           ||  | |||||||  |   | |||     ||  |     | |          |     | |
Db     285 GATGAACAGGATAGA---AGACTTGCAGAGCTTACCTCACAGAGAGCTTTAGAAACAGCT 341

Qy     280 ACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTAC 339
           || || |  ||||    ||      ||||     |   |   |    | |  |
Db     342 CCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAACGGCAATTACGGTTACGTAACA 401

Qy     340 CAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCC 399
           |    | | |   |      |  ||  |   || || ||||| || |     || | ||
Db     402 CCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGCCGTGAGTACACTCAAACAGTT 461

Qy     400 ATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGC 459
             | || |  ||  |  ||||  ||   ||||| |  || ||||  || || || || |
```

```
Db    462 GTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCATGCCGCCAACCTGACGGACAA 521

Qy    460 TGGCAGGTGATTA 472
          ||||| ||  | |
Db    522 TGGCAAGTTGTGA 534


RESULT 10
RIRANT17KD
LOCUS       RIRANT17KD               539 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.rickettsia 17K genus-common antigen gene, complete cds.
ACCESSION   M28479
VERSION     M28479.1  GI:152463
KEYWORDS    antigen.
SOURCE      R.rickettsia DNA.
  ORGANISM  Rickettsia rickettsii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Anderson,B.E. and Tzianabos,T.
  TITLE     Comparative sequence analysis of a genus-common rickettsial antigen
            gene
  JOURNAL   J. Bacteriol. 171 (9), 5199-5201 (1989)
  MEDLINE   89359171
   PUBMED   2768201
COMMENT     Draft entry and printed copy of sequence for [1] kindly submitted
            by B.Anderson, 27-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Rickettsia rickettsii"
                     /db_xref="taxon:783"
     -35_signal      1. .6
     -10_signal      21. .27
     mRNA            37. .>539
                     /product="17 kD mRNA"
     CDS             60. .539
                     /note="17 kD antigen precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26379.1"
                     /db_xref="GI:152464"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGKGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGQWQVVN"
     sig_peptide     60. .119
                     /note="17 kD antigen signal peptide"
     mat_peptide     120. .536
                     /product="17 kD antigen"
BASE COUNT      179 a     92 c    125 g    143 t
ORIGIN

  Query Match             8.3%;  Score 40.2;  DB 1;  Length 539;
  Best Local Similarity   46.1%;  Pred. No. 23;
  Matches  172;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy    100 GGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCG 159
          || |  | ||  ||||   |    ||    | |||||||| |||   || |   |
Db    165 GGCGGTGCTGGCGGCGCATTACTTGGTTCTCAATTCGGTAAGGGCAAAGGACAGCTTGTT 224

Qy    160 ATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATG 219
            |     | |||||  | || || |   |  ||||     ||||||||    | |||
Db    225 GGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTATG 284

Qy    220 GACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTT 279
          ||  | |||||||  |   | |||     ||  |     | |          |     | |
Db    285 GATGAACAGGATAGA---AGACTTGCAGAGCTTACCTCACAGAGAGCTTTAGAAACAGCT 341

Qy    280 ACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTAC 339
           || || |  ||||    ||     ||||     |    |   |    | |  |
Db    342 CCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAACGGCAATTACGGTTACGTAACA 401

Qy    340 CAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCC 399
          |   | | |   |     |  ||  |   || || ||||| || |     || | ||

Db    402 CCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGCCGTGAGTACACTCAAACAGTT 461

Qy    400 ATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGC 459
           | || |  ||  |  |||| ||   ||||| |  || ||||  || || || || |
Db    462 GTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCATGCCGCCAACCTGACGGACAA 521

Qy    460 TGGCAGGTGATTA 472
          ||||| ||  | |
Db    522 TGGCAAGTTGTGA 534


RESULT 11
AE008675
LOCUS       AE008675               10127 bp    DNA     linear   BCT 14-SEP-2001
DEFINITION  Rickettsia conorii Malish 7, section 107 of 114 of the complete
            genome.
ACCESSION   AE008675 AE006914
VERSION     AE008675.1  GI:15620413
KEYWORDS    .
SOURCE      Rickettsia conorii.
  ORGANISM  Rickettsia conorii
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (sites)
  AUTHORS   Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E.,
            Raoult,D. and Claverie,J.M.
  TITLE     Selfish DNA in protein-coding genes of Rickettsia
  JOURNAL   Science 290 (5490), 347-350 (2000)
  MEDLINE   20485642
   PUBMED   11030655
REFERENCE   2  (sites)
  AUTHORS   Ogata,H., Audic,S. and Claverie,J.-M.
  TITLE     Selfish DNA and the origin of genes
  JOURNAL   Science 291 (5502), 252-253 (2001)
REFERENCE   3  (bases 1 to 10127)
  AUTHORS   Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
            Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
            Raoult,D.
  TITLE     Mechanisms of evolution in Rickettsia conorii and R. prowazekii
  JOURNAL   Science 293 (5537), 2093-2098 (2001)
  MEDLINE   21442074
   PUBMED   11557893
REFERENCE   4  (bases 1 to 10127)
  AUTHORS   Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
            Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
            Raoult,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
            CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean
            Moulin, Marseille Cedex 05 13385, France
COMMENT     A public version of R. conorii genome database is accessible at
            http://igs-server.cnrs-mrs.fr/. The database intends to provide
            updated data. Annotation of the genome is an ongoing task whose
            goal is to make the genome sequence more useful. Comments to the
            authors are appreciated.
FEATURES             Location/Qualifiers
     source          1. .10127
                     /organism="Rickettsia conorii"
                     /strain="Malish 7"
                     /db_xref="taxon:781"
     gene            111. .890
                     /gene="map"
                     /note="RC1276"
     CDS             111. .890
                     /gene="map"
                     /codon_start=1
                     /transl_table=11
                     /product="methionine aminopeptidase [EC:3.4.11.18]"
                     /protein_id="AAL03814.1"
                     /db_xref="GI:15620414"
                     /translation="MTIKIHTGKDFIKMRAAGKLAAETLDFITDHVKPNVTTNSLNGL
                     CHNFITSHNAVPAPLNYKGFPKSICTSINHVVCHGIPNDKPLKNGDIVNIDVTVILDG
                     WYGDTSRMYYVGDVAIKPKRLIQVTYDAMMKGIEVVRPGAKLGDIGHAIQSYAEKHNY
```

```
                SVVRDYTGHGIGRVFHDEPSILNYGRSGTGLTLEEGMFFTVEPMINAGNYDTILSKLD
                GWTVTTRDKSLSAQFEHTIGVTKDGFEIFTLSPKKLDYPPY"
repeat_region   complement(977. .1011)
                /note="REP03, repeated element"
repeat_region   complement(1064. .1083)
                /note="REP02, RS3-like repeat"
gene            1141. .1452
                /gene="RC1277"
CDS             1141. .1452
                /gene="RC1277"
                /codon_start=1
                /transl_table=11
                /product="similarity to DNA repair protein (RadC)"
                /protein_id="AAL03815.1"
                /db_xref="GI:15620415"
                /translation="MLENNLLEFSNDEIIDTPHYIGHRKRVKERFVAAGAEYFSDYEL
                LEVMLFSAIPRKDIKPLAKKLLDHFDITDLINLDKERLLSIKGTNENLYINFALRELI
                N"
gene            1457. .1825
                /gene="RC1278"
CDS             1457. .1825
                /gene="RC1278"
                /codon_start=1
                /transl_table=11
                /product="similarity to DNA repair protein (RadC)"
                /protein_id="AAL03816.1"
                /db_xref="GI:15620416"
                /translation="MKQKIINKNINASWSSWIDYLKVNMGNMRLEQFRILFLNKKNIL
                IADEVLSQGTIDQAAVYPREIIKRALFNEASSLILVHHHPSGSPEPSKANIHMTNKIV
                EKCQTINIIVLNLICSYEYK"
gene            1812. .2795
                /gene="mraY2"
                /note="RC1279"
CDS             1812. .2795
                /gene="mraY2"
                /codon_start=1
                /transl_table=11
                /product="undecaprenyl-phosphate
                alpha-N-acetylglucosaminyltransferase [EC:2.4.1.-]"
                /protein_id="AAL03817.1"
                /db_xref="GI:15620417"
                /translation="MNINSYSTLFIFSFISTAILTYILTKYLPTIGLVDIPSNRRSHD
                KITPRGGGLAIVFVVMIALSGFEYITSNNFVNSIKILPLLLAIASISFLDDLKAVPIL
                IRLIVHLICAAFAIFLFSQLNSVHILIYSILVIALSGFINIYNFMDGIDGMSCVESIH
                LSSTMLILCFLQFPTINNPYFITSVNVIILGGSCGFLIFNWHPAKIFLGDVGSISLGF
                LTGLCLLLLALTSTNLFIACAIASLYYITDAVLTILIRLLNKEKIWQPHLKHFFQKAV
                QKGKSHKQVVSIIAVCNIFLMIISVTSLHFPVISIMLALAVITLTVQSLLK"
gene            2792. .3115
                /gene="RC1280"
CDS             2792. .3115
                /gene="RC1280"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAL03818.1"
                /db_xref="GI:15620418"
                /translation="MKLIKEDIDKIVRHIFAKQHPLLPEIMINWNKIVGFNFSTKALP
                LKITTYTYKKQKINTLFIQAEDNATAAELPYYQDIILERIKIYLGFAAIHQMNVTFYK
                GKKSL"
repeat_region   complement(3198. .3227)
                /note="REP02, RS3-like repeat"
repeat_region   3313. .3353
                /note="REP01, repeated element"
gene            complement(3509. .4258)
                /gene="RC1281"
CDS             complement(3509. .4258)
                /gene="RC1281"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAL03819.1"
                /db_xref="GI:15620419"
                /translation="MKKLLLIAAASTALLTSGLSFADCDMNSSVDSSTNSSMSSSVEN
                QWYLKLNAGGVIFNKTKPKGADFKLNNIKSNIKSNTGFTGEIGAGYYIMDNLRTDLTI
                GTVASSHLKKSKTYPDGNSFSVKNKPTIVSVLLNGYVDFVDLSMFKVFAGAGVGAAFV
                KEKIHSKDIKGGVTDTFNGTTKNKTNFAYQLSLGTSFEVAQGVKAELVYSWRDYGKTK
                NTTKTINGDKVKFGGTHYKGHNLMAGLRFDM"
repeat_region   4438. .4463
                /note="REP03, repeated element"
gene            complement(4505. .5188)
                /gene="RC1282"
CDS             complement(4505. .5188)
                /gene="RC1282"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAL03820.1"
                /db_xref="GI:15620420"
                /translation="MLIMKKLLLIAATSATMLSSSVSFAEGMDHEWYLRIDTGAAMFN
                EEKDKATGVKLKSNTTVPVALGIGYYISENFRADLTLGTIIGGKLKKSGAATNAPFTR
                TNISASHKPTITRLLINGYVDLTNFDMFDVFAGAGVGSALVKEKITYNGITGLSSNTK
                NRTNISYKLTLGTSAQIADGVKVELAYSWIDDGRTKSKNVIYQGTSVPTGGMHYQSHN
                LTAGIRFDI"
repeat_region   complement(5338. .5508)
                /note="REP06, A and T rich repeated element"
gene            complement(5478. .5828)
                /gene="fdxA"
                /note="RC1283"
CDS             complement(5478. .5828)
                /gene="fdxA"
                /codon_start=1
                /transl_table=11
                /product="ferredoxin"
                /protein_id="AAL03821.1"
                /db_xref="GI:15620421"
                /translation="MTYVVTDECVKCKYTDCVEVCPVDCFYEGEREDDFMLVINPDEC
                IDCGVCVPDCPIGAIKPESPGLIEWVERAKDFIENKGWKNITKKKTALPDADKFKDEK
                DKFNKYIGSVKIAN"
gene            complement(5833. .6525)
                /gene="ccmC"
                /note="RC1284"
CDS             complement(5833. .6525)
                /gene="ccmC"
                /codon_start=1
                /transl_table=11
                /product="heme exporter protein C"
                /protein_id="AAL03822.1"
                /db_xref="GI:15620422"
                /translation="MLKLLTPYYFNKLSKVIIPILAISTCVMMIIGLYLALIVSPVDY
                QQGDFVRIMYVHVPASWMALGVYVFMAVCSFSYLVWKTTISYLLAVASSYVGATFTLI
                SLVTGSLWGKPIWGTWWVWDARLTSMLILFLLYLSYIIIVNSADNIRKAQNPSSIIAL
                IGLINIPIVKFSVNIWYSLHQPASVLRLGSPTIHSSMLKPLIIMFISFILYFLLVLLL
                RTAILIDKIKNR"
gene            complement(6540. .7046)
                /gene="RC1285"
CDS             complement(6540. .7046)
                /gene="RC1285"
                /codon_start=1
                /transl_table=11
                /product="cytosine deaminase [EC:3.5.4.1]"
                /protein_id="AAL03823.1"
                /db_xref="GI:15620423"
                /translation="MNILLDKEANFNNFFMEQALKQAKIAFDKNEVPVGAVVVDRLHQ
                KIIASTHNNTEEKNNALYHAEIIAINEACNLISSKNLNDYDIYVTLEPCAMCAAAIAH
                SRLKRLFYGASDSKHGVVESNLRYFNSSACFHRPEIYSGILAEDSGLLMKEFFKRIRT
                VISSHRMT"
gene            complement(7027. .7947)
                /gene="p34"
                /note="RC1286"
CDS             complement(7027. .7947)
```

```
 Query Match             8.3%;  Score 40.2;  DB 1;  Length 10127;
 Best Local Similarity   46.1%;  Pred. No. 23;
 Matches  172;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy    100 GGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCG 159
```

```
Db    8239 GGCGGTGCTGGCGGCGCATTACTTGGTTCTCAATTCGGTAAGGGCAAAGGACAGCTTGTT 8298

Qy     160 ATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATG 219

Db    8299 GGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTATG 8358

Qy     220 GACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTT 279

Db    8359 GATGAACAGGATAGA---AGACTTGCAGAGCTTACCTCACAGAGAGCTTTAGAAACAGCT 8415

Qy     280 ACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTAC 339

Db    8416 CCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAACGGCAATTACGGTTACGTAACA 8475

Qy     340 CAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCC 399

Db    8476 CCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGCCGTGAGTACACTCAAACAGTT 8535

Qy     400 ATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGC 459

Db    8536 GTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCATGCCGCCAACCTGACGGACAA 8595

Qy     460 TGGCAGGTGATTA 472

Db    8596 TGGCAAGTTGTGA 8608


RESULT 12
AE004537
LOCUS       AE004537               13946 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 98 of 529 of the complete
            genome.
ACCESSION   AE004537 AE004091
VERSION     AE004537.1  GI:9946960
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 13946)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
   PUBMED   10984043
REFERENCE   2  (bases 1 to 13946)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .13946
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            69. .1247
                     /gene="PA1047"
     CDS             69. .1247
                     /gene="PA1047"
                     /codon_start=1
                     /transl_table=11
                     /product="probable esterase"
                     /protein_id="AAG04436.1"
                     /db_xref="GI:9946961"
                     /translation="MPRPLSHRNLPMTPQGHCDSRFAPLAEAFARLFDNPQERGAALC
                     LEVGGETVVDLWGGVADKDGEQPWQRDTILNLFSCTKTFTAVAVLQLVAEGKLELDAP
                     VAHYWPEFAAAGKAAISVRQLLCHRAGLPALREQMPPEALYDWQAMTTALAAEEPWWT
                     PGEAHGYAPITYGWLLGEVIRRVDGREPGEAIVARTAAPLGLDFHVGLDDSQFGRVAH
                     MARSKGSLGDAAAQRMLKTMMSEPLALTTRAFTNPPSILTSTNKPEWRRMQQPAANGH
                     GNARALAGFYSGLLQGRLLEDELLGQLAHEHAVGEDRTLLTRTRFGLGTMLDQPGVAN
                     ATYGLGPKAFGHPGAGGSIGFADPERDLAFSFVSNALGPYVLMDPRAQQLARVAGECL
                     "
     gene            1453. .2334
                     /gene="PA1048"
     CDS             1453. .2334
                     /gene="PA1048"
                     /codon_start=1
                     /transl_table=11
                     /product="probable outer membrane protein"
                     /protein_id="AAG04437.1"
                     /db_xref="GI:9946962"
                     /translation="MKLLKTSTLLLCMTASGCSYFQTQDASSQPAKAPAEQASSHSWW
                     PFGKDDGQTVAKQATEQKAEAEHVGSSHWWWPFGKDEGGRDVAAKAAEQKAEIAAAKK
                     EVAAAAEPKAANPWWWPFGASSASKASGKPALVQQKVTKEWLDAHEKALREAIAGSEF
                     TLERRDNALIVIAPADYSFNPKRHTMLMPITLNPLGKVAKMAQADPQSGILVLGHTDS
                     TGSKAVNDKLSFERAQSVAAIFRLSGLKGDQLRLKGVGSDMPRADNASAAGRAQNRRV
                     EVLYTQRASLLSLAQGN"
     gene            2433. .3080
                     /gene="pdxH"
                     /note="PA1049"
     CDS             2433. .3080
                     /gene="pdxH"
                     /codon_start=1
                     /transl_table=11
                     /product="pyridoxine 5'-phosphate oxidase"
                     /protein_id="AAG04438.1"
                     /db_xref="GI:9946963"
                     /translation="MTQSLADMRREYTRDGLSEANAPSDPFSLFRQWFDDAVKTERLP
                     VEPNAMTLATVDADGYPHCRILLLKGLDERGFTFFTNYESAKGRQLAANPRAAMTFFW
                     PALERQVRIEGSVEKVTPEESDAYYQVRPLGSRLGAWASPQSRVIADRAELERLLADT
                     ERRFADQPPSCPEHWGGYRLLPQRIEFWQGRPSRLHDRLDYRRQDGGWLRERLAP"
     gene            3088. .4200
                     /gene="PA1050"
     CDS             3088. .4200
                     /gene="PA1050"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG04439.1"
                     /db_xref="GI:9946964"
                     /translation="MFELDSTLAQHIVDRAMAILPHNINVMDAQGMIIGSGDPSRLHT
                     RHEGAQLVLANRRVVEIDEQAAACLRGVRPGVNLPLLHAERLVGVLGITGAPEVVRPY
                     AELVRMAAEMLMEQRQMQDERHWQRQRHEAWLRQLLDPGHGLGALAADAEHLGLAMHW
                     PRQAVFLELPEHADPLPCQARLLAALGGRAEHWSAALGERLVFWCRPVAHADTLEHWL
                     ERADARGWGVARLCAGDPAHDLAELRASACAARDLLDYARSRLPGRRVVTLDGHRLAS
                     LLFSQRDSWQVRQLLAPLQRLLDHDPQGSLAATLESWCAHDGHAQACAQALGIHRNSL
                     RYRLERIASISGLDLARLEHRQQLLLGLQLLVPRRG"
     gene            4359. .5717
                     /gene="PA1051"
     CDS             4359. .5717
                     /gene="PA1051"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transporter"
                     /protein_id="AAG04440.1"
                     /db_xref="GI:9946965"
                     /translation="MTLVLILLALIAFIVLSTTRLKLHPFLALLAAALIAGFAYQVPT
                     AEIVKTVTGGFGSILGYIGIVIVLGTIIGVILERSGAAITMAESVIRLLGERFPTLTM
                     SIIGYLVSIPVFCDSGFVILNSLKNALAARMKISTIAMSVALATGLYATHTFVPPTPG
                     PIAAAGNLGLDASLGLVIVVGLVVAFVTAMAGMWWANRFVGKDIPLVDDGQVVQTEED
                     FSELRARYGKLPSATQAFAPIFVPILLICLGSVAVFPSKPLGEGVLFACLNFLGQPVI
                     ALLVGLALACTLLKGQSKREEFHERVVEGIQSAAPILLITGAGGAFGAMLKITPLGDY
                     LGSTLSALGIGLFMPFVVAAALKTAQGSTTVALVTTSALVAPLLPQLGLDSEMGRVLT
```

```
                VMAIGAGAMTVSHANDSFFWVVTQFSRMTVSTAYRAQTLATLVQGIAGMLAVWLLSLV
                LL"
gene            5759. .6904
                /gene="PA1052"
CDS             5759. .6904
                /gene="PA1052"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG04441.1"
                /db_xref="GI:9946966"
                /translation="MKIVIAPDSFKESLSAPDVAEAIARGWRRVFPQAEVLLRPLADG
                GEGTVDAVLAATAGERRECRVEGPLGEPTLAHWGWLDDATAVIEMASASGLHLVPRDR
                RDATRSSSRGTGELIRAALDAGARKIILGLGGSATNDAGAGLLGALGVRFLAADGEEL
                APGGAALAGLHSLDLGGLDPRLVDVAVEVAADVDNPLCGPRGASAVFGPQKGASAEQV
                AQLDAALAHFAKVVAATLGEDFSRVPGVGAAGGLGFAARAFLRARFRPGIELVAELAG
                LADALVGADLVLTGEGRLDAQSLHGKTPVGVARIARQAGVPVVALAGSLGDGYQRLHD
                AGIDAAFSLAPGPISLEQACAGAAAELEARAEQIARLWRLAQASREG"
gene            7313. .7777
                /gene="PA1053"
CDS             7313. .7777
                /gene="PA1053"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG04442.1"
                /db_xref="GI:9946967"
                /translation="MRKSALIVASFTAMALALGGCQSSLTGDTYSREEARTVQTVRMG
                TIQALRPVKIEGTKTPIGSIAGAGVGGVAGSAVGGGKGSYVAAIIGAVAGGLLGAATE
                EGLTRTQGVEITVREDDGSTRAYVQQVDQGQIFRVGERVRILTVNGTSRVAH"
gene            8106. .10907
                /gene="PA1054"
CDS             8106. .10907
                /gene="PA1054"
                /codon_start=1
                /transl_table=11
                /product="probable NADH dehydrogenase"
                /protein_id="AAG04443.1"
                /db_xref="GI:9946968"
                /translation="MILALIIALPFMGALLPLLAERGGRTLCAASAGVAPLIGLILLL
                GQRGTVFAGQTLTASYSWLPEIGINLSLRLDGLGFLFALLILGIGLLVILYARYYLSE
                QEPMGRFFAFLLLFMGSMLGVVLAENLLLMMLFWELTSLSSFLLIGFWGSRSDARKGA
                RMALTVTGGGGLALLAGVLLLGHIAGSFELTEVLAAGELIRAHALYPLVLILILLGVF
                TKSAQFPFHFWLPHAMAAPTPVSAYLHSATMVKAGVFLLARLYPALAGTEWFFYLVSL
                SGLVTLLLGAGMALFQHDLKGLLAYSTISHLGLITLLFGLDSPLSNVAAVFHIINHAT
                FKASLFMAAGIIDHETGSRDMRLINGMWKYLPHTAVLAIVASLAMAGVPLLNGFLSKE
                MFFGETLSQNLLGSFNWLVPAVATLAGVFSVAYSLRFIHDVFFNGEPARLPKFPPHEP
                PRYMKVPVEILVFLCLLVGMLPAYTVAPLLAAAASASLGGQLPEYSLAIWHGFNLPLL
                MSVVALVGGVLVYVLRKPLFNWYAGLPEVDAKLVFEQQVQRVVALAARLTAWLENGSL
                QRYLAWLLGAALVVVAVELAPLARLTGSRGLTPLDGITALGMLVLAFSGLATALFHRM
                RLVALLILSVAGLLVALAFARFSAPDLALTQLSVEVVTMVLLMLALYFLPSRTPAESS
                SLRGLRDVALAGGVGMLVALLAYAVLTRPYESIAGFFVENSVSGGGGYNVVNVILVDF
                RGFDTLGEISVLAIAGVGIYAMLAGFALHKPTCDPQGRNWAHAKHPLILETLSRVLLP
                LALLISVFIFLRGHNLPGGGFIAGLVTAIALVLQYIASGSQWVEQRLPLNYAAMAGAG
                VLIAGLTGLGSWLFGYPFLTSAFGHFELPLIGEIELATAMLFDLGVYLTVVGAALLIL
                SYLGKLSLMPEPVAQEEAL"
gene            10907. .11245
                /gene="PA1055"
CDS             10907. .11245
                /gene="PA1055"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG04444.1"
                /db_xref="GI:9946969"
                /translation="MEALFAITLGLLTASGVYLLLRARTFPVVLGLTLISYAVNLFLF
                AMGRLGGSVAVIGTPGEYGDPLPQALVLTAIVIGFAMTAFVVVLALRGLGDLGSDHVD
                GRPDEEEPRP"
gene            11242. .12741
                /gene="PA1056"
CDS             11242. .12741
                /gene="PA1056"
                /codon_start=1
                /transl_table=11
                /product="probable NADH dehydrogenase"
                /protein_id="AAG04445.1"
                /db_xref="GI:9946970"
                /translation="MSHWLILPILLPLFAGSLLLLPLAERWQRGLSLLAALALIPLSL
                LLIRTAASGDLSVYALGNWAAPFGIVLMLDRLAALMLLATAVLGSAALIYALRGEDRL
                GKHFHALFQFQLLGINGAFLTGDLFNLFVFFEILLIASYALLLHGGGAERVRSGLHYV
                ILNLVGSAFFLIAVGTLYLGLTGTLNMADMAQKIAMADAERAPLLAAAGLLLLVVFALK
                AALLPLYFWLPRAYATASAPVAALFAIMTKVGIYSILRVYTLVFGDAAGELANLAQAW
                LWPLALATLGLGAIGALAARTLQGLLAYLVVVSAGTLLAGVALGSERALAASLYYLLH
                STWIAGGLFLLADLVARQRGDKAGDLVQGPALQNPRLLGGAFFIGAIAVAGLPPLSGF

  Query Match             8.2%;  Score 39.6;  DB 1;  Length 13946;
  Best Local Similarity   56.0%;  Pred. No. 33;
  Matches   75;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy      86 TTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCT 145
           | ||  |   | |||||||| |  || |||||||| ||||||     | | ||    |||
Db    7494 TCGGTTCCATCGCCGGTGCCGGCGTCGGCGGTGTCGCCGGCAGCGCGGTGGGCGGCGGCA 7553

Qy     146 CTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAA 205
             ||  |          |||||||  ||| |||||    ||||| ||| | ||| |    |
Db    7554 AGGGCAGCTACGTCGCGGCCATCATCGGCGCGGTCGCTGGCGGCCTGCTGGGCGCGGCTA 7613

Qy     206 TCGGTCAGAGCATG 219
            ||   |  || ||
Db    7614 CCGAAGAAGGCCTG 7627


RESULT 13
ATF7K2
LOCUS       ATF7K2                106702 bp    DNA     linear   PLN 10-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F7K2 (ESSA
            project).
ACCESSION   AL033545
VERSION     AL033545.2  GI:5738368
KEYWORDS    .
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 16021)
  AUTHORS   Bevan,M., Robben,J., Grymonprez,B., Volckaert,G., Bancroft,I.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 15770 to 106702)
  AUTHORS   Bevan,M., Wedler,H., Wambutt,R., Bancroft,I., Mewes,H.W., Lemcke,K.
            and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 92830 to 92989)
  AUTHORS   Bevan,M., Volckaert,G., Grymonprez,B., Voet,M., Robben,J.,
            Bancroft,I., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 106702)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Aug 18, 1999 this sequence version replaced gi:3892698.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .106702
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
```

```
misc_feature    1. .15769
                /note="basepairs 1-15769 are not integral part of BAC
                clone F7K2"
gene            3. .821
                /gene="AT4g22380"
gene            complement(join(3. .98,594. .821))
                /gene="AT4g22380"
CDS             complement(join(<3. .98,594. .821))
                /gene="AT4g22380"
                /note="strong similarity to NHP2/RS6 FAMILY PROTEIN, Homo
                sapiens, PID:g4826860
                Contains Ribosomal protein L7Ae signature AA73-90
                contains EST gb:F13981"
                /codon_start=1
                /product="Ribosomal protein L7Ae-like (fragment)"
                /protein_id="CAB52812.1"
                /db_xref="GI:5738369"
                /translation="MTGEVVNPKAYPLADSQLSITIMDLVQQATNYKQLKKGANEATK
                TLNRGISEFVVMAADAEPLEILLHLPLLAEDKNVPYVFVPSKQALGRACGVTRPVIAC
                SVTSNE"
exon            complement(3. .98)
                /gene="AT4g22380"
                /number=1
intron          complement(99. .593)
                /gene="AT4g22380"
                /number=1
exon            complement(594. .821)
                /gene="AT4g22380"
                /number=2
gene            1789. .2973
                /gene="AT4g22390"
gene            complement(1789. .2973)
                /gene="AT4g22390"
CDS             complement(1789. .2973)
                /gene="AT4g22390"
                /note="similarity to various predicted proteins,
                Arabidopsis thaliana"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB52813.1"
                /db_xref="GI:5738370"
                /translation="MAECPTDLINEMFLRLRATTLVKCRVLSKPCFSLIDSPEFVSSH
                LRRRLETGEHLMILLRGPRLLRTVELDSPENVSDIPHPLQAGGFTEVFGSFNGVIGLC
                NSPVDLAIFNPSTRKIHRLPIEPIDFPERDITREYVFYGLGYDSVGDDFKVVRIVQCK
                LKEGKKKFPCPVEVKVFSLKKNSWKRVCLMFEFQILWISYYYHLLPRRGYGVVVNNHL
                HWILPRRQGVIAFNAIIKYDLASDDIGVLSFPQELYIEDNMDIGVLDGCVCLMCYDEY
                SHVDVWVLKEYEDYKSWTKLYRVPKPESVESVEFIRPLICSKDRSKILLEINNAANLM
                WFDLESQSLTTAGIECDSSFTADILVSSLVLGCKGDPTQAQRSKDQKMMPKSTKRWYA
                LL"
exon            complement(1789. .2973)
                /gene="AT4g22390"
                /number=1
gene            4728. .5711
                /gene="AT4g22400"
CDS             4728. .5711
                /gene="AT4g22400"
                /note="similarity to various predicted proteins,
                Arabidopsis thaliana
                contains EST gb:Z17595"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB52814.1"
                /db_xref="GI:5738371"
                /translation="MLSSHLSQICFSCRDPLGNLPADLIRKCTDLMDLNFAEGQRLRT
                LNKNWKLALPSFRKGAYREERPWLLYRERGSGETRFFDPVRERVHRGNDPRLADARFL
                GSTLGWLVMSESLDLNPRKTHQTFLYNPFISELQQLPELTLDSPYPTGVLRYGVLTGN
                PSDNNTYACLITDYLHEIGNRRDTYTLLIYYVAKTSGRWEQNWSVASMLVVGIFDTWY
                IESISPSPDKISIEIWPKRYDFTFQTQEWNVSNFIPVDPIQWDPFHDQSLYHVKHRLQ
                LPENPSTLAIIIGTSTDRTRVRNKIIGDQSDDDDPRASDHIIDGVWVEVRP"
exon            4728. .5711
                /gene="AT4g22400"
                /number=1
gene            6112. .8034
                /gene="AT4g22410"
gene            complement(join(6112. .6297,6675. .6812,6908. .7075,
                7168. .7353,7429. .7513,7775. .8034))
                /gene="AT4g22410"
CDS             complement(join(6112. .6297,6675. .6812,6908. .7075,
                7168. .7353,7429. .7513,7775. .8034))
                /gene="AT4g22410"
                /note="similarity to snRNP assembly factor like protein,
                Homo sapiens, PID:g5730025"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB52815.1"
                /db_xref="GI:5738372"
                /translation="MVEFQVLDFHFERFCSVSLSNLNVYACLVCGKYFQGRSQKSHAY
                THSLEAGHHVYINLLTEKVYCLPDSYEINDPSLDDIRHVLNPRFSRAQVNELDKNRQW
                SRALDGSDYLPGMVGLNNIQKTEFVNVTIQSLMRVTPLRNFFHIPENYQHCKSPLVHC
                FGELTRKIWHARNFKGQVSPHEFLQAVMKASKKRFRIGQQSDPVEFMSWLLNTLHMDL
                RTSKDASSIIHKCFQGELEVVKEFQGNENKEISRMSFLMLGLDLPPPPLFKDVMEKNI
                IPQVALFDLLKKFDGETVTEVVRPKLARMRYRVIKSPRYLMFHMVRFKKNNFFKEKNP
                TLGESV"
exon            complement(6112. .6297)
                /gene="AT4g22410"
                /number=1
intron          complement(6298. .6674)
                /gene="AT4g22410"
                /number=1
exon            complement(6675. .6812)
                /gene="AT4g22410"
                /number=2
intron          complement(6813. .6907)
                /gene="AT4g22410"
                /number=2
exon            complement(6908. .7075)
                /gene="AT4g22410"
                /number=3
intron          complement(7076. .7167)
                /gene="AT4g22410"
                /number=3
exon            complement(7168. .7353)
                /gene="AT4g22410"
                /number=4
intron          complement(7354. .7428)
                /gene="AT4g22410"
                /number=4
exon            complement(7429. .7513)
                /gene="AT4g22410"
                /number=5
intron          complement(7514. .7774)
                /gene="AT4g22410"
                /number=5
exon            complement(7775. .8034)
                /gene="AT4g22410"
                /number=6
LTR             8059. .8449
                /note="389bp LTR"
misc_feature    8450. .13234
                /note="LTR retrotransposon"
LTR             13234. .13624
                /note="389bp LTR"
gene            13513. .14331
                /gene="AT4g22420"
gene            complement(join(13513. .13618,13701. .14033,14216. .14331))
                /gene="AT4g22420"
CDS             complement(join(13513. .13618,13701. .14033,14216. .14331))
                /gene="AT4g22420"
                /note="similarity to predicted protein, Arabidopsis
                thaliana
                contains EST gb:T44427"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB52816.1"
                /db_xref="GI:5738373"
                /translation="MAECPTDLINEMFLRLRATTLKKCRVLSKPCFSLIDSPEKRVIE
```

```
                RSDSPPPPLGFNNHLLPLANAYDDDDEEEGNELKKSQARRNGVAKGEGNGNKVNGEAQ
                EEVDDEEDDDDDASKGRGKHSRHVEVRRDCPYLDTVNRQVIIIDQFLMLRVPLATMRK
                RMRTGGRKAKAMEKYLKKVLKKSG"
     exon       complement(13513. .13618)
                /gene="AT4g22420"
                /number=1
     intron     complement(13619. .13700)
                /gene="AT4g22420"
                /number=1
     exon       complement(13701. .14033)
                /gene="AT4g22420"
                /number=2

  Query Match             8.2%;  Score 39.4;  DB 8;  Length 106702;
  Best Local Similarity   59.3%;  Pred. No. 37;
  Matches   67;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps  0;

Qy      85 GTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGC 144
           ||||| |  | | |||||| ||  | || || ||| | ||     |||| || |  | |
Db   33426 GTTGGAGGTGTCTCCGGTGGTGTGTTTGGAGGAGTTCCTGGTGGTGTGTTTGGAAGTGTC 33485

Qy     145 TCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGG 197
           |||||| ||||||     |   |   |||||  ||  | || ||| |   |||
Db   33486 TCTGGTGGTGTGTTTGGAGGAGTTCCCGGTGGTGTGTTTGGAGGTGTTTCTGG 33538


RESULT 14
ATCHRIV57
LOCUS       ATCHRIV57             199577 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57.
ACCESSION   AL161557
VERSION     AL161557.2  GI:7269071
KEYWORDS    .
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 48809 to 64829)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 64578 to 155510)
  AUTHORS   Wedler,H., Wambutt,R., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 141638 to 141797)
  AUTHORS   Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 143670 to 199577)
  AUTHORS   Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 199577)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV56 at the 5' end and an
            overlap with ATCHRIV58 at the 3' end.
FEATURES            Location/Qualifiers
     source         1. .199577
                    /organism="Arabidopsis thaliana"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="4"
     gene           6674. .9057
                    /gene="AT4g22260"
     gene           complement(join(6674. .6882,7268. .7339,7507. .7575,
                    7667. .7775,7863. .7978,8361. .8454,8588. .8777,8909. .9057))
                    /gene="AT4g22260"
     CDS            complement(join(6674. .6882,7268. .7339,7507. .7575,
                    7667. .7775,7863. .7978,8361. .8454,8588. .8777,8909. .9057))
                    /gene="AT4g22260"
                    /note="similarity to alternative oxidase, Mangifera
                    indica, PIR2:S45035
                    contains EST gb:T42793, AI994896.1, AI997645.1, R30022,
                    AA395166"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB79181.1"
                    /db_xref="GI:7269072"
                    /translation="MAAISGISSGTLTISRPLVTLRRSRAAVSYSSSHRLLHHLPLSS
                    RRLLLRNNHRVQATILQDDEEKVVVEESFKAETSTGTEPLEEPNMSSSSTSAFETWII
                    KLEQGVNVFLTDSVIKILDTLYRDRTYARFFVLETIARVPYFAFMSVLHMYETFGWWR
                    RADYLKVHFAESWNEMHHLLIMEELGGNSWWFDRFLAQHIATFYYFMTVFLYILSPRM
                    AYHFSECVESHAYETYDKFLKASGEELKNMPAPDIAVKYYTGGDLYLFENLYDVFVNI
                    RDDEAEHCKTMRACQTLGSLRSPHSILEDDDTEEESGCVVPEEAHCEGIVDCLKKSIT
                    S"
     exon           complement(6674. .6882)
                    /gene="AT4g22260"
                    /number=1
     intron         complement(6883. .7267)
                    /gene="AT4g22260"
                    /number=1
     exon           complement(7268. .7339)
                    /gene="AT4g22260"
                    /number=2
     intron         complement(7340. .7506)
                    /gene="AT4g22260"
                    /number=2
     exon           complement(7507. .7575)
                    /gene="AT4g22260"
                    /number=3
     intron         complement(7576. .7666)
                    /gene="AT4g22260"
                    /number=3
     exon           complement(7667. .7775)
                    /gene="AT4g22260"
                    /number=4
     intron         complement(7776. .7862)
                    /gene="AT4g22260"
                    /number=4
     exon           complement(7863. .7978)
                    /gene="AT4g22260"
                    /number=5
     intron         complement(7979. .8360)
                    /gene="AT4g22260"
                    /number=5
     exon           complement(8361. .8454)
                    /gene="AT4g22260"
                    /number=6
     intron         complement(8455. .8587)
                    /gene="AT4g22260"
                    /number=6
     exon           complement(8588. .8777)
                    /gene="AT4g22260"
                    /number=7
     intron         complement(8778. .8908)
                    /gene="AT4g22260"
                    /number=7
     exon           complement(8909. .9057)
                    /gene="AT4g22260"
                    /number=8
     gene           10103. .12393
                    /gene="AT4g22270"
     CDS            join(10103. .10244,10395. .10516,11327. .11779,11868. .12174,
                    12254. .12393)
                    /gene="AT4g22270"
```

```
                /note="contains EST gb:AI997575.1"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB79182.1"
                /db_xref="GI:7269073"
                /translation="MATNGSSEHLLPETTRPSAIINKQQPEFPASRFTFMSLVLWFDQ
                SNFGISFVSLSIWSRKFGMRRFLFLDKLWDVSDKVRIEYEAEIQRSLKRLMIFVLPSL
                TLEATYRIWWYISGFNQIPYIINPILSHVVACTLQLSSWLYRNSLFIIVCILYKITCH
                LQTLRLDDFARCFASEITDVRSALGEHQKIRRNLRIVSHRFRRFILLSLILVTATQFM
                ALLTTTRASVAVNIYEVGELALCSLSLVTGVFICLRSATKITHKAQSVTSLAAKWNVC
                ATVDSFDHLDGETPTGSIIESQVSLRGNAIETSDDEEGEGDDDLDNTKIHPIYANTIS
                YQKRQALVTYLENNKAGITVYGFLVDRSWLNTIFGIELALLLWLLNKTIGILA"
exon            10103. .10244
                /gene="AT4g22270"
                /number=1
intron          10245. .10394
                /gene="AT4g22270"
                /number=1
exon            10395. .10516
                /gene="AT4g22270"
                /number=2
intron          10517. .11326
                /gene="AT4g22270"
                /number=2
tRNA            11168. .11253
                /gene="AT4g22270"
                /product="tRNA-Tyr"
                /note="codon recognized: GUA"
exon            11327. .11779
                /gene="AT4g22270"
                /number=3
intron          11780. .11867
                /gene="AT4g22270"
                /number=3
exon            11868. .12174
                /gene="AT4g22270"
                /number=4
intron          12175. .12253
                /gene="AT4g22270"
                /number=4
exon            12254. .12393
                /gene="AT4g22270"
                /number=5
gene            12761. .15111
                /gene="AT4g22280"
gene            complement(join(12761. .13015,13081. .13227,13379. .14157,
                14209. .14314,14995. .15111))
                /gene="AT4g22280"
CDS             complement(join(12761. .13015,13081. .13227,13379. .14157,
                14209. .14314,14995. .15111))
                /gene="AT4g22280"
                /note="similarity to hypothetical protein, chr.4,
                Arabidopsis thaliana, PATCHX:G2392763
                contains EST gb:AI998861.1, AI993240.1"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB79183.1"
                /db_xref="GI:7269074"
                /translation="MEKDVMTDETCIVGRNPATIVNVNGLNISHQCMISLKLKCVIIF
                FGVVDAGDLLFRLITPERFIRIWQRMDRINEATSTSVLSMRWRYLFAFRPNLCLDDQE
                VGGGDSFIDFVDRVLVVTGNFPIRRISIKCRMSIDTGHVTRWMVDVLEHGVSYLDIDI
                ISEDDIGFVPLEIFTCKTLVELKLARGFYAMVPDYVSLPSLKTLFLSSIWFCNRDCCP
                LGRLLSACPVLEELTIIGGCWQHIEFCRTVSSSTLKKLTITSSFHYEYWGITIDTPSL
                AYLEYSDLVPRHFPFVNLESLVEAKLALDLCSDSNPTNLIKGLRNVEVLELSSGYTSQ
                ILYDFREAIPVLSKLFRLSIKSNSRHYYIWNLLPILLEKSPKLETLVIKGPLSADRYE
                REYGLSCPVKVLEITEYGGKYEELEQMEHFLKKLPCLELVKVRASAINDKEKSRITKD
                LLMVPRSSNCNIKLKFC"
exon            complement(12761. .13015)
                /gene="AT4g22280"
                /number=1
intron          complement(13016. .13080)
                /gene="AT4g22280"
                /number=1
exon            complement(13081. .13227)
                /gene="AT4g22280"
                /number=2
intron          complement(13228. .13378)
                /gene="AT4g22280"
                /number=2
exon            complement(13379. .14157)
                /gene="AT4g22280"
                /number=3
```

```
  Query Match              8.2%;  Score 39.4;  DB 8;  Length 199577;
  Best Local Similarity   59.3%;  Pred. No. 37;
  Matches   67;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy      85 GTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGC 144
           ||||| |  | | ||||||  ||  | || || ||| | ||     |||| || |  | |
Db   82234 GTTGGAGGTGTCTCCGGTGGTGTGTTTGGAGGAGTTCCTGGTGGTGTGTTTGGAAGTGTC 82293

Qy     145 TCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGG 197
           |||||| ||||||     |   |   |||||  ||  | || ||| |   |||
Db   82294 TCTGGTGGTGTGTTTGGAGGAGTTCCCGGTGGTGTGTTTGGAGGTGTTTCTGG 82346


RESULT 15
DMU60844/c
LOCUS       DMU60844                1070 bp    DNA     linear   INV 04-JUN-1998
DEFINITION  Drosophila melanogaster strain DPF-46 triose phosphate isomerase
            (Tpi) gene, complete cds.
ACCESSION   U60844
VERSION     U60844.1  GI:3184329
KEYWORDS    .
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1070)
  AUTHORS   Hasson,E., Wang,I.N., Zeng,L.W., Kreitman,M. and Eanes,W.F.
  TITLE     Nucleotide variation in the triosephosphate isomerase (Tpi) locus
            of Drosophila melanogaster and Drosophila simulans
  JOURNAL   Mol. Biol. Evol. 15 (6), 756-769 (1998)
  MEDLINE   98277691
   PUBMED   9615457
REFERENCE   2  (bases 1 to 1070)
  AUTHORS   Hasson,E., Wang,I.W., Zeng,L.W., Kreitman,M. and Eanes,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1996) Ecology & Evolution, State University of
            New York at Stony Brook, Stony Brook, NY 11794, USA
FEATURES             Location/Qualifiers
     source          1. .1070
                     /organism="Drosophila melanogaster"
                     /strain="DPF-46"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="99 E"
                     /note="individual isolate from strain DPF-46 (New York,
                     USA); Tpi slow allele"
     gene            <134. .>939
                     /gene="Tpi"
     mRNA            join(<134. .670,733. .>939)
                     /gene="Tpi"
                     /product="triose phosphate isomerase"
     CDS             join(134. .670,733. .939)
                     /gene="Tpi"
                     /codon_start=1
                     /product="triose phosphate isomerase"
                     /protein_id="AAC39049.1"
                     /db_xref="GI:3184330"
                     /translation="MSRKFCVGGNWKMNGDQKSIAEIAKTLSSAALDPNTEVVIGCPA
                     IYLMYARNLLPCELGLAGQNAYKVAKGAFTGEISPAMLKDIGADWVILGHSERRAIFG
                     ESDALIAEKAEHALAEGLKVIACIGETLEEREAGKTNEVVARQMCAYAQKIKDWKNVV
                     VAYEPVWAIGTGKTATPDQAQEVHAFLRQWLSDNISKEVSASLRIQYGGSVTAANAKE
                     LAKKPDIDGFLVGGASLKPEFVDIINARQ"
```

```
BASE COUNT      238 a    345 c    301 g    186 t
ORIGIN

  Query Match             8.1%;  Score 39;  DB 3;  Length 1070;
  Best Local Similarity   46.5%;  Pred. No. 46;
  Matches  126;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy      52 GTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTG 111
           || | |   | |||  |  || | |  |||| ||  | | | | |   |  ||  | |
Db     927 GTTGATGATGTCCACGAACTCGGGCTTCAGGGAGGCGCCTCCGACCAGGAAGCCATCGAT 868

Qy     112 GGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGC 171
           | |||  ||   ||||||||  | ||    |||    |||     || |  |   | |
Db     867 GTCGGGCTTCTTGGCCAGCTCCTTGGCGTTGGCGGCAGTCACAGATCCACCGTACTGGAT 808

Qy     172 GGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGAT 231
           |    || |   ||    ||  ||||   |    ||| ||||    |||  ||| | |
Db     807 GCGCAGGCTGGCGGACACCTCCTTGGAGATGTTGTCGCTCAGCCACTGGCGCAGGAAGGC 748

Qy     232 AAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGT 291
             | |    ||| || |    ||| ||  |  ||   | |    |  |    || ||
Db     747 GTGAACCTCTTGAGCCTGCAATTGGGAAGGGAAAATATCAGTTCTGGAAGGAATTAGCCA 688

Qy     292 AATCCGGACACCGGTAACAGCTACTCTGTGG 322
             | ||| |  |  || | | |    |||||
Db     687 TCTGCGGGCGACCTTACCTGATCGGGTGTGG 657


Search completed: February 22, 2003, 16:17:27
Job time : 1851.95 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:26:45 ; Search time 963.886 Seconds
(without alignments)
8165.910 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagaggatgtttgca.....aagtcatttcaacagaaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		53.8	11.1	998	17	BH133989	BH133989 ENTNW17TF
c 2		52.4	10.8	868	17	BH162606	BH162606 ENTRV08TF
c 3		49.6	10.2	279	17	CNS03W9F	AL263436 Tetraodon
c 4		47.2	9.7	845	17	AZ685166	AZ685166 ENTLG11TF
c 5		46.4	9.5	325	17	CNS03PK8	AL254753 Tetraodon
c 6		46.4	9.5	624	13	BJ001798	BJ001798 BJ001798
c 7		46.4	9.5	872	13	BJ009355	BJ009355 BJ009355
8		46	9.5	909	17	CNS00JTL	AL076720 Drosophil
c 9		46	9.5	939	17	CNS00CNG	AL059400 Drosophil
10		44.6	9.2	890	17	BH164851	BH164851 ENTQT04TF
c 11		43.2	8.9	427	9	AU052962	AU052962 AU052962
c 12		43.2	8.9	440	9	AU052941	AU052941 AU052941
c 13		43.2	8.9	715	13	BJ406237	BJ406237 BJ406237
c 14		42.8	8.8	630	9	AL677875	AL677875 AL677875
c 15		42.6	8.8	442	12	BE801262	BE801262 sr11g06.y
c 16		42.6	8.8	452	13	BI471035	BI471035 sah92f06.
c 17		42.6	8.8	536	17	CNS03Q0K	AL255341 Tetraodon
c 18		42.2	8.7	748	13	BI176637	BI176637 EST517582
c 19		42.2	8.7	805	17	AZ753057	AZ753057 RPCI-24-8
20		41.8	8.6	562	13	BJ382436	BJ382436 BJ382436
c 21		41.8	8.6	723	17	AZ627893	AZ627893 1M0469I18
c 22		41.8	8.6	745	17	CNS01MHE	AL150931 Anopheles
c 23		41.6	8.6	294	9	AU060898	AU060898 AU060898
c 24		41.2	8.5	1101	17	CNS00LO0	AL068607 Drosophil
c 25		41	8.4	328	9	AI437626	AI437626 sa37e06.y
c 26		41	8.4	344	12	BG352395	BG352395 sac03d01.
c 27		41	8.4	353	9	AI443457	AI443457 sa33h07.y
c 28		41	8.4	402	9	AI442486	AI442486 sa27f06.y
c 29		41	8.4	407	14	BQ742064	BQ742064 saq22d02.
c 30		41	8.4	431	10	BE555999	BE555999 sp96d11.y
c 31		41	8.4	460	14	BQ299690	BQ299690 sao42h05.
c 32		41	8.4	465	13	BM524492	BM524492 sal16d03.
c 33		41	8.4	467	14	BQ080883	BQ080883 san11e01.
c 34		41	8.4	485	13	BM524782	BM524782 sal19g02.
c 35		41	8.4	496	12	BE805704	BE805704 ss48f09.y
c 36		41	8.4	509	10	AW734942	AW734942 sk93b01.y
c 37		41	8.4	511	13	BM524641	BM524641 sal18a09.
c 38		41	8.4	520	13	BM093415	BM093415 saj08g01.
c 39		41	8.4	521	12	BG790034	BG790034 sae62c03.
c 40		41	8.4	536	13	BI784659	BI784659 saf57f12.
c 41		41	8.4	543	13	BI701733	BI701733 sa119e04.
c 42		41	8.4	569	14	BQ629443	BQ629443 saq02b10.
c 43		41	8.4	595	14	BQ473338	BQ473338 sap11d06.
44		40.8	8.4	805	17	CNS04RW2	AL304427 Tetraodon
c 45		40.6	8.4	551	14	BM959137	BM959137 PLATE_4_D

ALIGNMENTS

RESULT 1
BH133989
LOCUS BH133989 998 bp DNA linear GSS 07-AUG-2001
DEFINITION ENTNW17TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION BH133989
VERSION BH133989.1 GI:15093050
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 998)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 490.

```
FEATURES             Location/Qualifiers
     source          1. .998
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      348 a    144 c    238 g    268 t
ORIGIN

  Query Match             11.1%;  Score 53.8;  DB 17;  Length 998;
  Best Local Similarity   52.4%;  Pred. No. 2.9e-05;
  Matches  118;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy      24 TAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCA 83
           | || ||    | |||    |||   ||   | ||   |||        | || || ||
Db     321 TGGTGGTGCTGTCATTGGTGGTGCTGTTGGTGCTGTTGGTGGTGTTGTCATTGGTGGTGT 380

Qy      84 AGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAA 143
            |   | || | ||    ||| ||||||||||| ||||||| |||     |   ||||
Db     381 TGTCATTGGTGGTGTTGGTGGTGCTGTTGTTGGTGGTGTTGGTGGTGCTGTTGGTGGTGC 440

Qy     144 AGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTC 203
           | | |||||   ||||     |   ||| ||||| |||| | |||||  |  || | |
Db     441 TGTTGGTGGTGCTGTTGTTGGTGGTGTTGTTGGTGGTGTTGTTGGTGGTGTTGTTTGGTG 500

Qy     204 TAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAA 248
           ||     | | ||  |||  ||   |     || | |  |  |||
Db     501 TACTGATGATAAAAAGATAAATACCCCCACAAACAAACTGAAAAA 545


RESULT 2
BH162606/c
LOCUS       BH162606                868 bp    DNA     linear   GSS 24-SEP-2001
DEFINITION  ENTRV08TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH162606
VERSION     BH162606.1  GI:15736044
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 868)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 30
            High quality sequence stop: 768.
FEATURES             Location/Qualifiers
     source          1. .868
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      275 a    211 c     57 g    325 t
ORIGIN

  Query Match             10.8%;  Score 52.4;  DB 17;  Length 868;
  Best Local Similarity   54.7%;  Pred. No. 7.4e-05;
  Matches  104;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy      24 TAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCA 83
           | || ||    | |||    |||   ||   | ||   |||        | || || ||
Db     292 TGGTGGTGCTGTCATTGGTGGTGCTGTTGGTGCTGTTGGTGGTGTTGTCATTGGTGGTGT 233

Qy      84 AGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAA 143
            |   | || | ||    ||| ||||||||||| ||||||| |||     |   ||||
Db     232 TGTCATTGGTGGTGTTGGTGGTGCTGTTGTTGGTGGTGTTGGTGGTGCTGTTGGTGGTGC 173

Qy     144 AGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTC 203
           | | |||||   ||||     |   ||| ||||| |||| | |||||  |  |||||
Db     172 TGTTGGTGGTGCTGTTGTTGGTGGTGTTGTTGGTGGTGTTGTTGGTGGTGTTGTTGGTGG 113

Qy     204 TAAAATCGGT 213
           |    | |||
Db     112 TGTTGTTGGT 103


RESULT 3
CNS03W9F/c
LOCUS       CNS03W9F                279 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            063D14 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL263436
VERSION     AL263436.1  GI:7985098
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 279)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 279)
  AUTHORS   Genoscope.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .279
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="063D14"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG063DB07SP1~end :
                     PUC-Ori"
BASE COUNT      106 a    155 c      2 g     13 t      3 others
ORIGIN

  Query Match             10.2%;  Score 49.6;  DB 17;  Length 279;
  Best Local Similarity   56.8%;  Pred. No. 0.00033;
  Matches   88;  Conservative    1;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      47 TGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGG 106
           || |  |  ||||  |||   |     || || ||   |: || || | ||    ||| |
Db     219 TGGTGGTGATTGGTAGTGGTGATGGTGTTGGTGGTGNTGRTGTTGGTGGTGTTGGTGGTG 160

Qy     107 CTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGG 166
            || || ||  |||| || |||     |   |||||| ||| ||||| | ||||     |
Db     159 TTGGTGGTGCTGGTGGTGGTGGTGGTGTTGGTGGTAATGGTGGTGGTGGTGTTGGTGGTG 100

Qy     167 CCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           ||||||||| || | | |||||  |   ||||
Db      99 ATGGTGGTGGTGGTGGTGTTGGTGGTGTTGGTGGT 65


RESULT 4
AZ685166/c
LOCUS       AZ685166                845 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  ENTLG11TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ685166
VERSION     AZ685166.1  GI:11822312
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 26
            High quality sequence stop: 726.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                     method for isolate identification. Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library construction is described in detail in Smith,
                     H.O. and Venter, J.C. (Making small insert libraries for
                     whole genome shotgun sequencing projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      324 a    161 c     95 g    265 t
ORIGIN

  Query Match             9.7%;  Score 47.2;  DB 17;  Length 845;
  Best Local Similarity   52.6%;  Pred. No. 0.0029;
  Matches  103;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      74 TTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGC 133
           || ||| |   |   | || | ||    ||| | ||||| ||| | ||||| |||
Db     743 TTGGTCCTGTTGTCATTGGTGGTGTTGTTGGTGGTGTTGATGGTGCTGTTGGTGGTGCTG 684

Qy     134 TGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGAT 193
           |   ||||   | | |||||   ||||     ||  ||||||||||||  | |||||
Db     683 TTGGTGGTGCTGTTGGTGGTGATGTTGGTGGTGCTGTTGGTGGTGCTGTCATTGGTGGTG 624

Qy     194 TAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGA 253
           |  ||||| |   | |||          | |  | |  | | ||  ||    ||  | |
Db     623 TTGTTGGTGGTGTTGTTGGTGTTAGTGATGTTATGAAAAAGACAAGTAACACCAAAAAAA 564

Qy     254 GTTTGGAAAAGGTAAA 269
           |    |  |  ||||
Db     563 TTGAAAATTAATTAAA 548


RESULT 5
CNS03PK8/c
LOCUS       CNS03PK8                325 bp    DNA     linear   GSS 17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            045I18 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL254753
VERSION     AL254753.1  GI:7975765
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 325)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 325)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Tetraodon nigroviridis"
```

```
                     /db_xref="taxon:99883"
                     /clone="045I18"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG045BE09SP1~end :
                     PUC-Ori"
BASE COUNT      157 a    157 c     10 g      0 t      1 others
ORIGIN

  Query Match             9.5%;  Score 46.4;  DB 17;  Length 325;
  Best Local Similarity   52.6%;  Pred. No. 0.0034;
  Matches  101;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy      22 GGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGT 81
           ||| || || | || || |  ||||   | | |||||   ||         ||  |  |
Db     237 GGTGGTGGTGTTGTTGTTGTTGGTGTTGGTGTTGTTGGTGCTGGTGGTGGTGTTGTTGTT 178

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
             |  |  |  | |||   ||| | || |||||  | ||||| ||      |  |||||
Db     177 GTTGGTGGTGTTGGTGCTGGTGGTGGTGGTGTTGTTGTTGTTGGTGTTGGTGTTGTTGGT 118

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           ||| ||||| | ||||   | |    || ||||| ||||   | |||      ||||
Db     117 GCTGGTGGTGGTGGTGTTGTTGTTGTTGGTGTTGGTGTTGTTGGTGCTGGTGGTGGTGGT 58

Qy     202 TCTAAAATCGGT 213
             |    | |||
Db      57 GTTGTTGTTGGT 46


RESULT 6
BJ001798/c
LOCUS       BJ001798                624 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ001798 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA027D02 5',
            mRNA sequence.
ACCESSION   BJ001798
VERSION     BJ001798.1  GI:17364689
KEYWORDS    EST.
SOURCE      Japanese medaka.
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Oryzias latipes"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA027D02"
                     /clone_lib="MF01SSA cDNA"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
BASE COUNT      208 a    248 c     64 g    104 t
ORIGIN

  Query Match             9.5%;  Score 46.4;  DB 13;  Length 624;
  Best Local Similarity   52.0%;  Pred. No. 0.0045;
  Matches  104;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      14 GTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACT 73
           ||||  | |||  | ||   | | |   |  ||  |||   | ||     | | |    |
Db     453 GTTTCGATGGTTTTGGTGGTACTGTAGACGATGGTTTTGGTGGTGCTGTAGACGATGGTT 394

Qy      74 TTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGC 133
           || || ||   | || | | | ||| | ||| |||||  |  |  ||| ||| |||
Db     393 TTGGTGGTGCTGTAGACTGTGGTGCTAATGGTGCTGTAGACGATGGTTTTGGTGGTGCTG 334

Qy     134 TGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGAT 193
           |    |||     | |||||   || |      |   |||||| | ||||    |||||
Db     333 TAGACGGTGGTTTTGGTGGTGCTGTAGACGGTGGTTTTGGTGATACTGTAGACGGTGGTG 274

Qy     194 TAATTGGTTCTAAAATCGGT 213
               |||| ||| |  ||||
Db     273 CTGGTGGTACTATAGACGGT 254


RESULT 7
BJ009355/c
LOCUS       BJ009355                872 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ009355 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA131D04 5',
            mRNA sequence.
ACCESSION   BJ009355
VERSION     BJ009355.1  GI:17356777
KEYWORDS    EST.
SOURCE      Japanese medaka.
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Oryzias latipes"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA131D04"
                     /clone_lib="MF01SSA cDNA"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
BASE COUNT      250 a    318 c    130 g    169 t      5 others
ORIGIN

  Query Match             9.5%;  Score 46.4;  DB 13;  Length 872;
  Best Local Similarity   52.0%;  Pred. No. 0.0052;
  Matches  104;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      14 GTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACT 73
           ||||  | |||  | ||   | | |   |  ||  |||   | ||     | | |    |
Db     762 GTTTCGATGGTTTTGGTGGTACTGTAGACGATGGTTTTGGTGGTGCTGTAGACGATGGTT 703

Qy      74 TTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGC 133
           || || ||   | || | | | ||| | ||| |||||  |  |  ||| ||| |||
Db     702 TTGGTGGTGCTGTAGACTGTGGTGCTAATGGTGCTGTAGACGATGGTTTTGGTGGTGCTG 643

Qy     134 TGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGAT 193
           |    |||     | |||||   || |      |   |||||| | ||||    |||||
Db     642 TAGACGGTGGTTTTGGTGGTGCTGTAGACGGTGGTTTTGGTGATACTGTAGACGGTGGTG 583

Qy     194 TAATTGGTTCTAAAATCGGT 213
               |||| ||| |  ||||
Db     582 CTGGTGGTACTATAGACGGT 563
```

```
RESULT 8
CNS00JTL
LOCUS       CNS00JTL                909 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR39G08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL076720
VERSION     AL076720.1  GI:4956298
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .909
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR39G08"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      293 a     55 c    175 g    144 t    242 others
ORIGIN

  Query Match             9.5%;  Score 46;  DB 17;  Length 909;
  Best Local Similarity   26.7%;  Pred. No. 0.007;
  Matches   52;  Conservative   66;  Mismatches   77;  Indels    0;  Gaps    0;

Qy      10 GGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAG 69
           || :|: :|   :|  |:  | |  :: |: |  |::|:|||:| :||:| :|:|
Db     685 GGGKGKGKGKKKKGGKGKGTTKTTKKKKTKKTGKGKKTKTTTKTTKTTKGGKGKGGGGGT 744

Qy      70 AACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGC 129
             |:: || ::   :  :: ::   |: |  || |: | ::::|:| :|:  :: :|:
Db     745 KGKTKKTGTTKKTKTKKKKKNKKKTKTKTGTKTGMGKKTTKKKKTKGGKGKTGKKGKGKG 804

Qy     130 CAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGT 189
             | :|:::|::   :|: ::||:   :  |    ||    :|| : :: :::|:::::|
Db     805 KGGKKGKKKGKKGKGKGKKKKGGKGYGKGGGKGKGGGGGGGKGGGKTKKTKKKTKKKKKT 864

Qy     190 GGATTAATTGGTTCT 204
           :  :|  |:: || |
Db     865 KTGKTKTTKKTTTKT 879


RESULT 9
CNS00CNG/c
LOCUS       CNS00CNG                939 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059400
VERSION     AL059400.1  GI:4946964
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR26H16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       71 a    349 c    104 g    180 t    235 others
ORIGIN

  Query Match             9.5%;  Score 46;  DB 17;  Length 939;
  Best Local Similarity   6.4%;  Pred. No. 0.0071;
  Matches   14;  Conservative   133;  Mismatches   71;  Indels    0;  Gaps    0;

Qy       7 AGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCC 66
           |:| : :::: :   :::::: :: :  ::::: :  ::::::::::|::: : ::::
Db     572 AKAAKKKKKKGKKKKKKDDKKKKKKKKKKKDKKKKAKKKKKKKKKKKAKKKAKKKKKKK 513

Qy      67 CAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCT 126
             : :  :::::: ::   ::  :: :: :  : :  ::: : :::::: :: :::::: :
Db     512 DDADKKKKKDKKKKKAKDDKKKKKKDKKDAKKKKKKKKAKKKKKKKKAKKKKKKKKKK 453

Qy     127 GGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTG 186
           ::     :::::::::| |:|:  : ::  |::::    ::   ::::::: : :::::::
Db     452 KKAKKAAKKKKKKKKAKAKGKKAKTDKKAAKKKKDKKAKKKKKKKKKKDKAKKKKKKKK 393

Qy     187 GGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGA 224
           ::::: :  |::|::: :      |:: ||: : ::||
Db     392 KKKKKKKGGAKDGKKKGKGGGGGGGKKRAADRKKKKGA 355


RESULT 10
BH164851
LOCUS       BH164851                890 bp    DNA     linear   GSS 24-SEP-2001
DEFINITION  ENTQT04TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH164851
VERSION     BH164851.1  GI:15738289
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
```

```
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 633.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      334 a     65 c    212 g    279 t
ORIGIN

  Query Match             9.2%;  Score 44.6;  DB 17;  Length 890;
  Best Local Similarity   65.7%;  Pred. No. 0.019;
  Matches   65;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     103 GGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCA 162
           | || ||||||||| ||||||| |||     |   ||||    || |||||   ||||
Db     596 GTGGGTGTTGTTGGTGGTGTTGATGGTGCTGTTGGTGGTGCTCGTGGTGGTGCTGTTGGT 655

Qy     163 ATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
             ||  |||||||||||||||   ||||   | ||||||
Db     656 GGTGCTGTTGGTGGTGCTGTTGGTGGTGCTGTCATTGGT 694


RESULT 11
AU052962/c
LOCUS       AU052962                427 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU052962 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLF389, mRNA sequence.
ACCESSION   AU052962
VERSION     AU052962.1  GI:4701445
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLF389"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      231 a    105 c     25 g     66 t
ORIGIN

  Query Match             8.9%;  Score 43.2;  DB 9;  Length 427;
  Best Local Similarity   51.6%;  Pred. No. 0.037;
  Matches   99;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      11 GATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGA 70
           |||| |    || || || | | ||| |||| | | |  || | ||||    ||
Db     357 GATGGTGCAGAGTTATTACTATTATTGTTATTATTATTATTATTGTTGTTGTTGTTGTTG 298

Qy      71 ACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCC 130
             ||  |  |   |  || |  | ||    ||  | ||||||||  | ||||| ||
Db     297 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 238

Qy     131 AGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTG 190
             |  ||| |   | |  || |   ||||   | |   ||| || || |||| | | ||
Db     237 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 178

Qy     191 GATTAATTGGTT 202
             |  ||| ||
Db     177 TTGTTGTTGTTT 166


RESULT 12
AU052941/c
LOCUS       AU052941                440 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU052941 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLF355, mRNA sequence.
ACCESSION   AU052941
VERSION     AU052941.1  GI:4701424
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLF355"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      234 a    110 c     26 g     70 t
ORIGIN

  Query Match             8.9%;  Score 43.2;  DB 9;  Length 440;
  Best Local Similarity   51.6%;  Pred. No. 0.037;
```

```
  Matches   99;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      11 GATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGA 70
           |||| |    || || || | | ||| |||| | | |  || | ||||    ||
Db     370 GATGGTGCAGAGTTATTACTATTATTGTTATTATTATTATTATTGTTGTTGTTGTTGTTG 311

Qy      71 ACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCC 130
             ||  |  |   |  || |  | ||    ||  | ||||||||  | ||||| ||
Db     310 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 251

Qy     131 AGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTG 190
             |  ||| |   | |  || |   ||||   | |   ||| || || |||| | | ||
Db     250 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 191

Qy     191 GATTAATTGGTT 202
             |  ||| ||
Db     190 TTGTTGTTGTTT 179


RESULT 13
BJ406237/c
LOCUS       BJ406237                715 bp    mRNA    linear   EST 10-MAR-2002
DEFINITION  BJ406237 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds34i16 3', mRNA sequence.
ACCESSION   BJ406237
VERSION     BJ406237.1  GI:19319154
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds34i16"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
                     /sex="mat A"
                     /dev_stage="Slug stage"
BASE COUNT      284 a    166 c     71 g    190 t      4 others
ORIGIN

  Query Match             8.9%;  Score 43.2;  DB 13;  Length 715;
  Best Local Similarity   53.6%;  Pred. No. 0.046;
  Matches   90;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      33 AATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGG 92
           | ||||| |  ||||  || | | ||    ||         ||  | ||  ||  || ||
Db     332 AGTTATTGTTGGTGTCGTTGTTGGTGTTGTTGTTGGTGTAGTTGTTGGTGTAGTTGTTGG 273

Qy      93 AGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGG 152
           ||  |    ||| |  |||||||| |  |||| |||     |  ||||   || |  |||
Db     272 AGTAGTTGTTGGAGTAGTTGTTGGAGTAGTTGTTGGAGTAGTTGTTGGAGTAGTTGTTGG 213

Qy     153 TCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGG 200
           |||||     |   ||| ||||| ||| | || |   |  ||||
Db     212 AGTAGTTGTTGGTGTAGTTGTTGGTGTGGTTGTTGGAGTGGTTGTTGG 165


RESULT 14
AL677875/c
LOCUS       AL677875                630 bp    mRNA    linear   EST 18-MAR-2002
DEFINITION  AL677875 XGC-neurula Silurana tropicalis cDNA clone TNeu058i22 5',
            mRNA sequence.
ACCESSION   AL677875
VERSION     AL677875.1  GI:19534249
KEYWORDS    EST.
SOURCE      western clawed frog.
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Taylor R
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TNeu058i22.p1cSP6
            Sequencing primer: P1CSP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Silurana tropicalis"
                     /db_xref="taxon:8364"
                     /clone="TNeu058i22"
                     /clone_lib="XGC-neurula"
                     /dev_stage="neurula"
                     /lab_host="Escherichia coli DH10B"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from neurula.
                     EcoRI-NotI cut cDNA was then ligated into pCS107 with
                     EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      204 a    136 c    117 g    173 t
ORIGIN

  Query Match             8.8%;  Score 42.8;  DB 9;  Length 630;
  Best Local Similarity   50.0%;  Pred. No. 0.058;
  Matches  107;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       4 AACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGT 63
           || | | | | | |||    |    || |   | || || ||||  | || | ||    |
Db     497 AAGAAACGCTTTGTGCTGCCTTAATGTATTTGTCTTTTCCGTGTTCTCTTGGATGTGGTT 438

Qy      64 GCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTT 123
             | ||| | | ||  | | || || |||||  |||| |  ||  | |  |  ||  | ||
Db     437 CCACAGTAATGTATCCATTAATAATTCGGATTTGCGGCGTGGTGTCTGTTGAGCTTTCTT 378

Qy     124 GCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTT 183
           |   || | |     | | |  | | |||    |||    |  || ||| |    || ||
Db     377 GAATGCGAACCCAGAGATCACTGAAATGGCGATGTTTTTCTTTCCTTTGCTTCCTCTTTT 318

Qy     184 TTGGGTGGATTAATTGGTTCTAAAATCGGTCAAT 217
           |    |   ||   |  || | ||   |||| ||
Db     317 TGTACTTTCTTTCGTCTTTATCAAGCTGGTCGAT 284


RESULT 15
BE801262/c
LOCUS       BE801262                442 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  sr11g06.y1 Gm-c1050 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1050-83 5' similar to TR:Q40786 Q40786 ARABINOGALACTAN-PROTEIN
            PRECURSOR. ;, mRNA sequence.
ACCESSION   BE801262
VERSION     BE801262.1  GI:10232374
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
```

```
            Glycine.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 893   Std Error: 0.00
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .442
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1050-83"
                     /clone_lib="Gm-c1050"
                     /tissue_type="leaf tissue at various developmental stages
                     of greenhouse grown plants"
                     /dev_stage="3 week old"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The Clark NIL was constructed and seed was provided
                     by Dr. J. Specht, University of Nebraska (Shoemaker and
                     Specht, 1995). The cDNA library was constructed from mRNA
                     isolated from leaf tissue at various  developmental stages
                     of 3 week old greenhouse grown plants. Complementary  DNA
                     was synthesized from mRNA using a primer consisting of a
                     poly(dT)  sequence with a XhoI restriction site and a 3'
                     anchor. EcoRI adapters  were ligated to the blunt-ended
                     cDNA fragments followed by XhoI digestion.  The cDNA
                     fragments were directionally cloned into the EcoRI-XhoI
                     restriction  site of the pBluescript vector. The ligated
                     cDNA fragments were transformed  into DH10B host cells
                     (GibcoBRL). The library was constructed in  cooperation
                     with Dr. Paul Keim's laboratory at Northern Arizona
                     University."
BASE COUNT      107 a    195 c     71 g     69 t
ORIGIN

  Query Match             8.8%;  Score 42.6;  DB 12;  Length 442;
  Best Local Similarity   53.3%;  Pred. No. 0.057;
  Matches   90;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy      35 TTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAG 94
           ||  | |  | ||   | | ||||   |||   |       | || |   ||  |  |  |
Db     236 TTGGTGTTGGCGTTGGTGTTGTTGCTGGTGGTGATGTTGCTGGTGGAGGAGTTGCTGCTG 177

Qy      95 CTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTC 154
           |||    ||| | || || ||| ||||||||||||  | ||  || |   |   ||| |
Db     176 CTGGTGGTGGTGTTGCTGGTGGTGGTGTTGCTGGCGGGGTGGGTGTTGCGTGCAGGTGCTG 117

Qy     155 GAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTC 203
             |||     |    || ||| | |||  ||||||| |   | |||||
Db     116 ATCTTGGTGGTGGTGGTGATGGGGTTGTAGTGGGTGGTTGTGTGGGTTC 68


Search completed: February 22, 2003, 17:05:09
Job time : 1028.89 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:36:28 ; Search time 40.2902 Seconds
(without alignments)
3699.288 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagaggatgtttgca.....aagtcatttcaacagaaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	47.8	9.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	42.6	8.8	465	4	US-08-476-102A-2	Sequence 2, Appli
	3	35.8	7.4	303	4	US-08-556-978B-80	Sequence 80, Appl
	4	35.8	7.4	303	4	US-08-556-978B-81	Sequence 81, Appl
	5	35.4	7.3	303	4	US-08-556-978B-82	Sequence 82, Appl
c	6	35.2	7.2	198	5	PCT-US95-10668-3	Sequence 3, Appli
c	7	35.2	7.2	198	5	PCT-US95-10668-4	Sequence 4, Appli
c	8	34.4	7.1	1046	1	US-08-361-467B-4	Sequence 4, Appli
c	9	34.4	7.1	1046	1	US-08-484-332C-4	Sequence 4, Appli
c	10	33.6	6.9	198	5	PCT-US95-10668-1	Sequence 1, Appli
c	11	33.6	6.9	198	5	PCT-US95-10668-2	Sequence 2, Appli
	12	33.2	6.8	1011	2	US-08-825-781-2	Sequence 2, Appli
	13	33.2	6.8	2072	2	US-09-073-362-2	Sequence 2, Appli
	14	33.2	6.8	2072	2	US-09-243-920-2	Sequence 2, Appli
	15	33	6.8	377	2	US-08-332-766A-1	Sequence 1, Appli
c	16	32.8	6.7	1690	1	US-08-276-452A-24	Sequence 24, Appl
c	17	32.8	6.7	1690	2	US-08-798-744-24	Sequence 24, Appl
c	18	32.4	6.7	1608	4	US-09-292-225-20	Sequence 20, Appl
	19	32.4	6.7	1608	4	US-09-292-225-22	Sequence 22, Appl
c	20	32.4	6.7	1665	4	US-09-292-225-17	Sequence 17, Appl
	21	32.4	6.7	1665	4	US-09-292-225-19	Sequence 19, Appl
c	22	32.4	6.7	1752	4	US-09-292-225-14	Sequence 14, Appl
	23	32.4	6.7	1752	4	US-09-292-225-16	Sequence 16, Appl
c	24	32.4	6.7	1984	1	US-07-885-970A-25	Sequence 25, Appl
c	25	32.4	6.7	1985	1	US-08-298-687A-25	Sequence 25, Appl
c	26	32.4	6.7	1985	1	US-08-298-829-25	Sequence 25, Appl
	27	32.2	6.6	1440	4	US-09-134-001C-1598	Sequence 1598, Ap
c	28	32.2	6.6	1561	1	US-07-968-971A-3	Sequence 3, Appli
c	29	32.2	6.6	1561	1	US-07-824-247-44	Sequence 44, Appl
c	30	32.2	6.6	1561	1	US-08-142-473A-4	Sequence 4, Appli
c	31	32.2	6.6	1561	1	US-08-469-203A-4	Sequence 4, Appli
c	32	32.2	6.6	1561	1	US-08-469-203A-4	Sequence 4, Appli
c	33	32.2	6.6	1561	3	US-08-470-204A-44	Sequence 44, Appl
c	34	32.2	6.6	1561	4	US-08-926-522-10	Sequence 10, Appl
c	35	31.8	6.5	379	1	US-08-145-617-5	Sequence 5, Appli
	36	31.8	6.5	2793	1	US-08-209-747-1	Sequence 1, Appli
	37	31.8	6.5	2793	1	US-08-458-298-1	Sequence 1, Appli
	38	31.8	6.5	5496	4	US-09-462-284-1	Sequence 1, Appli
c	39	31.6	6.5	1632	1	US-08-324-243-34	Sequence 34, Appl
c	40	31.6	6.5	1632	1	US-08-532-390-34	Sequence 34, Appl
c	41	31.6	6.5	1632	3	US-08-717-294-34	Sequence 34, Appl
c	42	31.6	6.5	1632	5	PCT-US95-11511-34	Sequence 34, Appl
c	43	31.6	6.5	3656	4	US-09-221-017B-786	Sequence 786, App
c	44	31.4	6.5	925	3	US-08-858-003-1	Sequence 1, Appli
c	45	31.4	6.5	925	3	US-09-078-166-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 9.8%; Score 47.8; DB 1; Length 7218;

```
  Best Local Similarity   2.0%;  Pred. No. 3.1e-05;
  Matches    7;  Conservative  206;  Mismatches  138;  Indels    0;  Gaps    0;

Qy     136 TTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTA 195
           ||||||| ::: :: ::  :::  : :: ::  :  :: :: :  :    ::: :::  :
Db    1442 TTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1383

Qy     196 ATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGT 255
           :  ::    ::::  ::  ::  :: :::  :: :::: :::: ::::  :::  ::::
Db    1382 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1323

Qy     256 TTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGT 315
             :::::::: ::::: :::: ::: :: : :  :: : ::   ::: : ::: :: ::
Db    1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1263

Qy     316 TATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAA 375
           : :: :  :::  :: : : :   :  :: :  : :: ::: ::::: :  :  :: ::
Db    1262 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1203

Qy     376 TATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGC 435
           :  :  :::::    :: :::::: :: ::  : :::: ::::: :::::::   : ::
Db    1202 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1143

Qy     436 ACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486
           :  : : :  :: ::  ::: ::  :  :: :::  :    :: :::::::
Db    1142 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092


RESULT 2
US-08-476-102A-2
; Sequence 2, Application US/08476102A
; Patent No. 6355450
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann, et al.
;        TITLE OF INVENTION: Nucleotide Sequence of the Haemphilus
; Patent No. 6355450
;                           influenzae Rd Genome, Fragments Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Ave.
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/476,102A
;             FILING DATE: 07-Jun-1995
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Marks, Michelle S.
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 465 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-476-102A-2

  Query Match             8.8%;  Score 42.6;  DB 4;  Length 465;
  Best Local Similarity   54.9%;  Pred. No. 0.00033;
  Matches   84;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy      75 TAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCT 134
           || ||      | ||| ||  |       ||| |  | | | || ||| |||||||
Db     171 TAATCAAGGTGTAGTTGGTACGCTTGGTGGTGGAGCTTTAGGTGGTATTGCTGGTAGTAC 230

Qy     135 GTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATT 194
             ||||    ||| |||||| || |       ||  | | |||||  ||   |||| | |
Db     231 AATTGGCGGTGGTCGTGGTCAAGCTATTGCAGCAGTAGTTGGTGCAATTGGCGGTGCAAT 290

Qy     195 AATTGGTTCTAAAATCGGTCAATCGATGGATCA 227
           |  |||   ||||||||   ||   |||  |||
Db     291 AGCTGGAAGTAAAATCGAAGAAAAAATGAGTCA 323


RESULT 3
US-08-556-978B-80
; Sequence 80, Application US/08556978B
; Patent No. 6268169
;  GENERAL INFORMATION:
;    APPLICANT:  FAHNESTOCK, STEPHEN F.
;    TITLE OF INVENTION:  NOVEL RECOMBINANTLY PRODUCED
;    TITLE OF INVENTION:  SPIDER SILK ANALOGS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,978B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/077,600
;      FILING DATE:  JUNE 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9389-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  80:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-556-978B-80

  Query Match             7.4%;  Score 35.8;  DB 4;  Length 303;
  Best Local Similarity   56.3%;  Pred. No. 0.04;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy      84 AGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAA 143
           ||  |  |  || ||   ||| |||| |   |||||| |||     ||   |   ||| |
Db     134 AGCCGCTGCGGCAGCTGGTGGTGCTGGTCAGGGCGGTCTTGGCTCACAAGGGGCCGGTCA 193

Qy     144 AGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTT 202
           |||   ||||  ||  |||   |||  ||| |||||  |     |||||||    ||| |
Db     194 AGGCGCTGGTGCAGCAGCAGCTGCCGCTGGCGGTGCAGGCCAAGGTGGATATGGTGGCT 252


RESULT 4
```

```
US-08-556-978B-81
; Sequence 81, Application US/08556978B
; Patent No. 6268169
;  GENERAL INFORMATION:
;    APPLICANT:  FAHNESTOCK, STEPHEN F.
;    TITLE OF INVENTION:  NOVEL RECOMBINANTLY PRODUCED
;    TITLE OF INVENTION:  SPIDER SILK ANALOGS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,978B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/077,600
;      FILING DATE:  JUNE 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9389-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  81:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-556-978B-81

  Query Match             7.4%;  Score 35.8;  DB 4;  Length 303;
  Best Local Similarity   56.3%;  Pred. No. 0.04;
  Matches   67;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy      84 AGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAA 143
           ||  |  |  || ||   ||| |||| |   |||||| |||     ||   |   ||| |
Db      83 AGCCGCTGCGGCAGCTGGTGGTGCTGGTCAGGGCGGTCTTGGCTCACAAGGGGCCGGTCA 142

Qy     144 AGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTT 202
           |||   ||||  ||  |||   |||  ||| ||||| |     |||||||    ||| |
Db     143 AGGCGCTGGTGCAGCAGCAGCTGCCGCTGGCGGTGCAGGCCAAGGTGGATATGGTGGCT 201


RESULT 5
US-08-556-978B-82
; Sequence 82, Application US/08556978B
; Patent No. 6268169
;  GENERAL INFORMATION:
;    APPLICANT:  FAHNESTOCK, STEPHEN F.
;    TITLE OF INVENTION:  NOVEL RECOMBINANTLY PRODUCED
;    TITLE OF INVENTION:  SPIDER SILK ANALOGS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,978B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/077,600
;      FILING DATE:  JUNE 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9389-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  82:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-556-978B-82

  Query Match             7.3%;  Score 35.4;  DB 4;  Length 303;
  Best Local Similarity   53.2%;  Pred. No. 0.054;
  Matches   75;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      62 GTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTG 121
           ||| ||||       ||  |   |  | |  |||||   ||| || | |   || |||
Db      68 GTGGCCAGGGTGCAGGTGCTGCTGCTGCTGCGGCTGCTGGTGGTGCAGGTCAGGGTGGTC 127

Qy     122 TTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTG 181
           | |      |||      ||| |||||  ||||  || |||   |||   |||||| || |
Db     128 TGGGATCTCAGGGCGCAGGTCAAGGTGCTGGTGCAGCTGCGGCGGCAGCTGGTGGCGCGG 187

Qy     182 TTTTGGGTGGATTAATTGGTT 202
           |   ||||| |     ||||
Db     188 GTCAAGGTGGCTACGGCGGTT 208


RESULT 6
PCT-US95-10668-3/c
; Sequence 3, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-3

  Query Match             7.2%;  Score 35.2;  DB 5;  Length 198;
  Best Local Similarity   49.5%;  Pred. No. 0.05;
  Matches   91;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      18 GCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAG 77
           ||  | |  |  | |  || || |   |||  || | ||||    ||         ||
Db     185 GCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 126

Qy      78 TCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTT 137
           |  |   |  || |  | ||    ||  | ||||||||  | |||||  ||      |  |
Db     125 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 66

Qy     138 TGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAAT 197
           || |   | |  || |   |||| |||     ||| || || |||| | | ||   |  |
Db      65 TGTTGTTGTTGTTGTTGTTGTTGGAATTCTTGTTGTTGTTGTTGTTGTTGTTGTTGT 6

Qy     198 TGGT 201
           || |
Db       5 TGTT 2


RESULT 7
PCT-US95-10668-4/c
; Sequence 4, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-4

  Query Match             7.2%;  Score 35.2;  DB 5;  Length 198;
  Best Local Similarity   49.5%;  Pred. No. 0.05;
  Matches   91;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      18 GCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAG 77
           ||  | |  |  | |  || || |   |||  || | ||||    ||         ||
Db     185 GCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 126

Qy      78 TCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTT 137
           |  |   |  || |  | ||    ||  | ||||||||  | |||||  ||      |  |
Db     125 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 66

Qy     138 TGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAAT 197
           || |   | |  || |   |||| |||     ||| || || |||| | | ||   |  |
Db      65 TGTTGTTGTTGTTGTTGTTGTTGGAATTCTTGTTGTTGTTGTTGTTGTTGTTGTTGT 6

Qy     198 TGGT 201
           || |
Db       5 TGTT 2


RESULT 8
US-08-361-467B-4/c
; Sequence 4, Application US/08361467B
; Patent No. 5633441
;  GENERAL INFORMATION:
;    APPLICANT:  De Greef, Willy
;    APPLICANT:  Van Emmelo, John
;    APPLICANT:  De Oliveria, Dulce E.
;    APPLICANT:  De Souza, Maria-Helena
;    APPLICANT:  Van Montagu, Marc
;    TITLE OF INVENTION:  PLANTS WITH MODIFIED FLOWERS, SEEDS OR
;    TITLE OF INVENTION:  EMBRYOS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,467B
;      FILING DATE:  22-DEC-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,492
;      FILING DATE:  04-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/EP90/01275
;      FILING DATE:  01-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 89 402 224.3
;      FILING DATE:  04-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schulman, Robert M.
;      REGISTRATION NUMBER:  31,196
;      REFERENCE/DOCKET NUMBER:  010830-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
```

```
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    IMMEDIATE SOURCE:
;      CLONE:  3C9
US-08-361-467B-4

  Query Match             7.1%;  Score 34.4;  DB 1;  Length 1046;
  Best Local Similarity   52.9%;  Pred. No. 0.22;
  Matches   74;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      84 AGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAA 143
           || ||  |||   |    ||| | || |||||| ||| |  ||||     |   ||||
Db     329 AGGAGCAGGATAAGGTGATGGTGTTGGTGTTGGTGGTTTAACTGGTGGCTTAGCTGGTGG 270

Qy     144 AGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTC 203
             |  ||| | |   |     ||| | | ||||| | |    ||||| |||| ||||
Db     269 TTTTGATGGGCTATAAGTGGGAGCCTTAGTTGGTGGTTTAACCGGTGGTTTAACTGGTGG 210

Qy     204 TAAAATCGGTCAATCGATGG 223
           |      || | ||   |||
Db     209 TTTCGATGGGCTATAAGTGG 190


RESULT 9
US-08-484-332C-4/c
; Sequence 4, Application US/08484332C
; Patent No. 5767374
;  GENERAL INFORMATION:
;    APPLICANT:  De Greef, Willy
;    APPLICANT:  Van Emmelo, John
;    APPLICANT:  De Oliveria, Dulce E.
;    APPLICANT:  De Souza, Maria-Helena
;    APPLICANT:  Van Montagu, Marc
;    TITLE OF INVENTION:  PLANTS WITH MODIFIED FLOWERS, SEEDS OR
;    TITLE OF INVENTION:  EMBRYOS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,332C
;      FILING DATE:  7-JUNE-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/361,467
;      FILING DATE:  22-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,492
;      FILING DATE:  04-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/EP90/01275
;      FILING DATE:  01-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 89 402 224.3
;      FILING DATE:  04-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schulman, Robert M.
;      REGISTRATION NUMBER:  31,196
;      REFERENCE/DOCKET NUMBER:  010830-093
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    IMMEDIATE SOURCE:
;      CLONE:  3C9
US-08-484-332C-4

  Query Match             7.1%;  Score 34.4;  DB 1;  Length 1046;
  Best Local Similarity   52.9%;  Pred. No. 0.22;
  Matches   74;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      84 AGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAA 143
           || ||  |||   |    ||| | || |||||| ||| |  ||||     |   ||||
Db     329 AGGAGCAGGATAAGGTGATGGTGTTGGTGTTGGTGGTTTAACTGGTGGCTTAGCTGGTGG 270

Qy     144 AGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTC 203
             |  ||| | |   |     ||| | | ||||| | |    ||||| |||| ||||
Db     269 TTTTGATGGGCTATAAGTGGGAGCCTTAGTTGGTGGTTTAACCGGTGGTTTAACTGGTGG 210

Qy     204 TAAAATCGGTCAATCGATGG 223
           |      || | ||   |||
Db     209 TTTCGATGGGCTATAAGTGG 190


RESULT 10
PCT-US95-10668-1/c
; Sequence 1, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
```

```
PCT-US95-10668-1

  Query Match             6.9%;  Score 33.6;  DB 5;  Length 198;
  Best Local Similarity   48.9%;  Pred. No. 0.16;
  Matches   90;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy      18 GCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAG 77
           ||  | |  |  | |  || || |   |||  || | ||||    ||          ||
Db     185 GCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 126

Qy      78 TCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTT 137
           |  |    |  || |  | ||    ||  | ||||||||  | |||||  ||      |  |
Db     125 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 66

Qy     138 TGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAAT 197
           || |    | |  || |   ||||  ||     ||| || || |||| | | ||   |  |
Db      65 TGTTGTTGTTGTTGTTGTTGTTGGGATCCTTGTTGTTGTTGTTGTTGTTGTTGTTGT 6

Qy     198 TGGT 201
           || |
Db       5 TGTT 2


RESULT 11
PCT-US95-10668-2/c
; Sequence 2, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-2

  Query Match             6.9%;  Score 33.6;  DB 5;  Length 198;
  Best Local Similarity   48.9%;  Pred. No. 0.16;
  Matches   90;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy      18 GCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAG 77
           ||  | |  |  | |  || || |   |||  || | ||||    ||          ||
Db     185 GCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 126

Qy      78 TCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTT 137
           |  |    |  || |  | ||    ||  | ||||||||  | |||||  ||      |  |
Db     125 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 66

Qy     138 TGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAAT 197
           || |    | |  || |   ||||  ||     ||| || || |||| | | ||   |  |
Db      65 TGTTGTTGTTGTTGTTGTTGTTGGGATCCTTGTTGTTGTTGTTGTTGTTGTTGTTGT 6

Qy     198 TGGT 201
           || |
Db       5 TGTT 2


RESULT 12
US-08-825-781-2
; Sequence 2, Application US/08825781
; Patent No. 5843727
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN TUMOR-ASSOCIATED MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/825,781
;      FILING DATE:  Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0262 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PANCTUT02
;      CLONE:  2236771
US-08-825-781-2

  Query Match             6.8%;  Score 33.2;  DB 2;  Length 1011;
  Best Local Similarity   54.0%;  Pred. No. 0.53;
  Matches   68;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy      97 GCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGA 156
           || | ||  ||||||| ||    | |||| |  ||| | ||||| | |  ||   |   |
Db     209 GCCAGTGATGCTGTTGCTGTGACTTTTGCAGATCAGATATTTGGAATATTTAACTGGATA 268

Qy     157 GTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAA 216
             || || || |  |||       |   ||| |||||  | | ||  || | |  | | |
Db     269 ATTCCACTGTCAGTTGCATTATCCTGTTTTGGTGGCCTCAATGCCTCCATTGTGGCTGCT 328
```

```
Qy     217 TCGATG 222
           || | |
Db     329 TCTAGG 334


RESULT 13
US-09-073-362-2
; Sequence 2, Application US/09073362
; Patent No. 5942399
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  AMINO ACID PERMEASE HOMOLOG
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/073,362
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0514 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2072 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  ESOGTUT02
;      CLONE:  2667831
US-09-073-362-2

  Query Match             6.8%;  Score 33.2;  DB 2;  Length 2072;
  Best Local Similarity   54.0%;  Pred. No. 0.77;
  Matches   68;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy      97 GCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGA 156
           || | ||  ||||||| ||    | |||| |  ||| | ||||| | |  ||   |   |
Db    1015 GCCAGTGATGCTGTTGCTGTGACTTTTGCAGATCAGATATTTGGAATATTTAACTGGATA 1074

Qy     157 GTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAA 216
           || || || |  |||       |   ||| |||||  | | ||  || |   | | |
Db    1075 ATTCCACTGTCAGTTGCATTATCCTGTTTTGGTGGCCTCAATGCCTCCATTGTGGCTGCT 1134

Qy     217 TCGATG 222
           || | |
Db    1135 TCTAGG 1140


RESULT 14
US-09-243-920-2
; Sequence 2, Application US/09243920
; Patent No. 5981242
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  AMINO ACID PERMEASE HOMOLOG
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/243,920
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/073,362
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0514 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2072 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  ESOGTUT02
;      CLONE:  2667831
US-09-243-920-2

  Query Match             6.8%;  Score 33.2;  DB 2;  Length 2072;
  Best Local Similarity   54.0%;  Pred. No. 0.77;
  Matches   68;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy      97 GCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGA 156
           || | ||  ||||||| ||    | |||| |  ||| | ||||| | |  ||   |   |
Db    1015 GCCAGTGATGCTGTTGCTGTGACTTTTGCAGATCAGATATTTGGAATATTTAACTGGATA 1074

Qy     157 GTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAA 216
           || || || |  |||       |   ||| |||||  | | ||  || |   | | |
Db    1075 ATTCCACTGTCAGTTGCATTATCCTGTTTTGGTGGCCTCAATGCCTCCATTGTGGCTGCT 1134

Qy     217 TCGATG 222
           || | |
Db    1135 TCTAGG 1140


RESULT 15
US-08-332-766A-1
; Sequence 1, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;       STREET:  1100 New York Avenue, N.W.
;       CITY:  Washington
;       STATE:  D. C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005-3918
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/332,766A
;       FILING DATE:  01-NOV-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9326052.9
;       FILING DATE:  21-DEC-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BIRD, Donald J.
;       REGISTRATION NUMBER:  25,323
;       REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 861-3000
;       TELEFAX:  (202) 822-0944
;       TELEX:  6714627 CUSH
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  377 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-1

  Query Match             6.8%;  Score 33;  DB 2;  Length 377;
  Best Local Similarity   53.5%;  Pred. No. 0.36;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy      78 TCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTT 137
           || ||  ||   |   | ||    ||  | || || ||| |||| || |||  |  |
Db      66 TCCTCCTGATAACAATGGTGATGATGATGGTGATGATGGTGGTGGTGATGGTGATGTTGG 125

Qy     138 TGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAAT 197
           |||| | | | ||| | | | ||    |    |||||||| || |   |||||
Db     126 TGGTGATGATGGTGATGGTGATGGTGGTGATGATGGTGGTGGTGATGGTGGTGGTGATGG 185

Qy     198 TGGTTCTAA 206
           ||||  | |
Db     186 TGGTGGTGA 194


Search completed: February 22, 2003, 17:07:09
Job time : 63.2902 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 16:18:29 ; Search time 47.285 Seconds
(without alignments)
5772.590 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagaggatgtttgca.....aagtcatttcaacagaaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	43.2	8.9	155074	9	US-10-026-188-6	Sequence 6, Appli
	2	39.6	8.1	522	10	US-09-864-761-19900	Sequence 19900, A
	3	39	8.0	350	10	US-09-864-761-23424	Sequence 23424, A
	4	39	8.0	600	10	US-09-864-761-6690	Sequence 6690, Ap
	5	38.8	8.0	887	10	US-09-864-761-21484	Sequence 21484, A
	6	38.8	8.0	1981	10	US-09-864-761-4745	Sequence 4745, Ap
	7	38.4	7.9	1671	10	US-09-815-242-6926	Sequence 6926, Ap
	8	38.2	7.9	446	10	US-09-864-761-20699	Sequence 20699, A
	9	37.8	7.8	1075	10	US-09-864-761-19241	Sequence 19241, A
	10	37.8	7.8	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	11	37.6	7.7	470	10	US-09-864-761-3121	Sequence 3121, Ap
	12	37.4	7.7	249	10	US-09-864-761-16846	Sequence 16846, A
	13	37.4	7.7	476	10	US-09-864-761-6	Sequence 6, Appli
c	14	36.8	7.6	1635	10	US-09-864-761-20241	Sequence 20241, A
c	15	36.8	7.6	1973	10	US-09-864-761-3471	Sequence 3471, Ap
c	16	36	7.4	253	10	US-09-864-761-20903	Sequence 20903, A
c	17	36	7.4	331	10	US-09-864-761-17053	Sequence 17053, A
c	18	36	7.4	454	10	US-09-864-761-4149	Sequence 4149, Ap
c	19	36	7.4	465	10	US-09-864-761-230	Sequence 230, App
c	20	35.6	7.3	257	10	US-09-878-574-10174	Sequence 10174, A
	21	35.6	7.3	456	10	US-09-864-761-4249	Sequence 4249, Ap
	22	35.2	7.2	401	10	US-09-864-761-3936	Sequence 3936, Ap
	23	35.2	7.2	454	10	US-09-881-752A-23	Sequence 23, Appl
	24	35.2	7.2	512	10	US-09-864-761-25347	Sequence 25347, A
	25	35.2	7.2	575	10	US-09-864-761-8628	Sequence 8628, Ap
	26	34.6	7.1	14776	10	US-09-880-107-2088	Sequence 2088, Ap
c	27	34.4	7.1	370	10	US-09-728-446-223	Sequence 223, App
c	28	34	7.0	462	10	US-09-864-761-2794	Sequence 2794, Ap
c	29	33.8	7.0	439	10	US-09-864-761-20174	Sequence 20174, A
	30	33.8	7.0	581	10	US-09-822-849A-31	Sequence 31, Appl
c	31	33.8	7.0	867	10	US-09-216-393-340	Sequence 340, App
c	32	33.8	7.0	867	10	US-09-216-393-342	Sequence 342, App
c	33	33.8	7.0	1397	10	US-09-216-393-343	Sequence 343, App
	34	33.8	7.0	1397	10	US-09-216-393-345	Sequence 345, App
	35	33.6	6.9	479	10	US-09-864-761-768	Sequence 768, App
	36	33.6	6.9	830	10	US-09-864-761-19531	Sequence 19531, A
	37	33.4	6.9	335913	9	US-09-754-853A-2	Sequence 2, Appli
	38	33.4	6.9	335913	9	US-09-754-853A-3	Sequence 3, Appli
	39	33.2	6.8	1582	9	US-10-163-866-12	Sequence 12, Appl
	40	33.2	6.8	2098	9	US-10-163-866-11	Sequence 11, Appl
c	41	33.2	6.8	2614	9	US-09-822-846-491	Sequence 491, App
c	42	33	6.8	549	9	US-10-114-893-41	Sequence 41, Appl
	43	33	6.8	4032	10	US-09-815-242-6940	Sequence 6940, Ap
c	44	32.8	6.7	276	10	US-09-864-761-25120	Sequence 25120, A
c	45	32.8	6.7	334	10	US-09-878-574-5330	Sequence 5330, Ap

ALIGNMENTS

RESULT 1
US-10-026-188-6
; Sequence 6, Application US/10026188
; Patent No. US20020164645A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Zhang, Yifeng
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
; TITLE OF INVENTION: Ion Channel
; FILE REFERENCE: 02307E-114910US
; CURRENT APPLICATION NUMBER: US/10/026,188
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/259,379
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 155074
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human genomic region containing ltrpc5
; OTHER INFORMATION: (Human Chromosome 11p15.5 PAC clone pDJ915f1
; OTHER INFORMATION: containing KvLQT1 gene)
US-10-026-188-6

Query Match 8.9%; Score 43.2; DB 9; Length 155074;
Best Local Similarity 51.6%; Pred. No. 0.023;
Matches 99; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy      22 GGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGT 81
           | | || ||     | ||    |||    | | | |||| |||   ||     | |  ||
Db  45577 GATGGTGGTGGTGGTTTTGGTGGTGGTGATGTTGGTGGCAGTGATGAGGGTGGTGGCAGT 45636

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
           | |  | | | | ||    ||| | || | |||| |||| || |||       |||||
Db  45637 GATGGTGACAGTGTTGATGGTGGTGGTGGTTTTGGTGGTGGTGATGGAGGCGGTGTTGGT 45696

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           ||| |||||   | ||     |    || ||||| || |   ||||| ||   ||||
Db  45697 GTTGGTGGTGGTTTTGGTGGCGGTGATGGTGATGGTGTTGGTGGTGGTGGTTTTGGTGGT 45756
```

```
Qy     202 TCTAAAATCGGT 213
           | |    |||
Db   45757 GGTGATGGTGGT 45768


RESULT 2
US-09-864-761-19900
; Sequence 19900, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19900
;   LENGTH: 522
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL049749.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.4
;   OTHER INFORMATION: EST_HUMAN HIT: AA776664.1, EVALUE 4.90e-01
US-09-864-761-19900

  Query Match             8.1%;  Score 39.6;  DB 10;  Length 522;
  Best Local Similarity   53.2%;  Pred. No. 0.014;
  Matches   84;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy      33 AATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGG 92
           | | || || | |||  | |  | |||  |||  | ||     ||  | |||| |  ||
Db     121 AGTGATCATGAAGGTGGTGATGATTGATGGTGGTAATAATGACTGTGATAAAGATGGTGG 180

Qy      93 AGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGG 152
             | ||  | ||| | || || ||| |||| || |||        || |   | |  |||
Db     181 TGATGAAAGTGGTGATGGTGATGGTGGTGATGGTGGTGGTGGTGATGATGGTGATGATGG 240

Qy     153 TCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTG 190
           | |   ||     | || || ||||| || |   ||||
Db     241 TGGTAGTGATGGTGGCAGTGATGGTGGTGGTAATGGTG 278


RESULT 3
US-09-864-761-23424
; Sequence 23424, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```
; SEQ ID NO 23424
;   LENGTH: 350
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019159.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
;   OTHER INFORMATION: NT HIT: AF111944.1, EVALUE 4.00e-03
;   OTHER INFORMATION: EST_HUMAN HIT: AA484400.1, EVALUE 3.20e-01
US-09-864-761-23424

  Query Match             8.0%;  Score 39;  DB 10;  Length 350;
  Best Local Similarity   50.8%;  Pred. No. 0.018;
  Matches   93;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      7 AGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCC 66
          || ||  || |  | ||| || ||     || | | | ||    |   | |||  ||||
Db     52 AGTGGTGGTATTGATGGTGGTGGTGGGCATAATGTTAATGCTGATGGTGATGGTGGTGCT 111

Qy     67 CAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCT 126
           |      | || || | | ||  |  | ||    ||  | || || ||| | || || |
Db    112 GATGGGGGTGGTGGTGATGGAGGTGATGATGATGGTGATGGTGGTGGTGGTGATGATGAT 171

Qy    127 GGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTG 186
          ||         ||| || ||| ||| |   | |      | | |||||||| ||    |
Db    172 GGTGGTGGTCGTGGCAATGGTGGTGATAATGGTTGTGGTGATAGTGGTGGTGGTGAGGGG 231

Qy    187 GGT 189
          |||
Db    232 GGT 234


RESULT 4
US-09-864-761-6690
; Sequence 6690, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6690
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC019159.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
US-09-864-761-6690

  Query Match             8.0%;  Score 39;  DB 10;  Length 600;
  Best Local Similarity   50.8%;  Pred. No. 0.024;
  Matches   93;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      7 AGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCC 66
          || ||  || |  | ||| || ||     || | | | ||    |   | |||  ||||
Db    216 AGTGGTGGTATTGATGGTGGTGGTGGGCATAATGTTAATGCTGATGGTGATGGTGGTGCT 275

Qy     67 CAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCT 126
           |      | || || | | ||  |  | ||    ||  | || || ||| | || || |
Db    276 GATGGGGGTGGTGGTGATGGAGGTGATGATGATGGTGATGGTGGTGGTGGTGATGATGAT 335

Qy    127 GGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTG 186
          ||         ||| || ||| ||| |   | |      |  | |||||||| ||    |
Db    336 GGTGGTGGTCGTGGCAATGGTGGTGATAATGGTTGTGGTGATAGTGGTGGTGGTGAGGGG 395

Qy    187 GGT 189
          |||
Db    396 GGT 398


RESULT 5
US-09-864-761-21484
; Sequence 21484, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
```

```
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21484
;   LENGTH: 887
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005414.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  11
;   OTHER INFORMATION: EST_HUMAN HIT: BE787945.1, EVALUE 3.00e-03
;   OTHER INFORMATION: NT HIT: S65019.1, EVALUE 1.20e-01
US-09-864-761-21484

  Query Match             8.0%;  Score 38.8;  DB 10;  Length 887;
  Best Local Similarity   50.0%;  Pred. No. 0.034;
  Matches   97;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67
           ||||  || |    ||| | | |     || |   || |    ||| | |||  ||
Db      84 GAGGTGGTGTTGGTGGTGGCAATGATGGTAATGGTAGAGGTAGTTTGGATGGTGGTAATG 143

Qy      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127
             | |   | ||  |   ||    |||| ||  |  || | || || ||| ||||  | ||
Db     144 GGGATGGTGGTAATGGTGATAGTGGAGATGGTAGAGGTGGTGATGGTGGTGGTGGGGATG 203

Qy     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187
           |  |      ||||   |||| |||| | | |     |    || || || || |   |
Db     204 GTGATGATGGTGGTGGTGGTAATGGTGGTGGTAATGATGGTGGTGATGATGGTGGTGATG 263

Qy     188 GTGGATTAATTGGT 201
           ||||    | ||||
Db     264 GTGGTGGTAATGGT 277
```

```
RESULT 6
US-09-864-761-4745
; Sequence 4745, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4745
;   LENGTH: 1981
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005414.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  11
US-09-864-761-4745

  Query Match             8.0%;  Score 38.8;  DB 10;  Length 1981;
```

```
  Best Local Similarity   50.0%;  Pred. No. 0.052;
  Matches   97;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67
           ||||  || |    ||| | | |     || |   || |    ||| | |||  ||
Db     291 GAGGTGGTGTTGGTGGTGGCAATGATGGTAATGGTAGAGGTAGTTTGGATGGTGGTAATG 350

Qy      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127
           | |   | ||  |   ||    |||| ||  |  || | || || ||| ||||  | ||
Db     351 GGGATGGTGGTAATGGTGATAGTGGAGATGGTAGAGGTGGTGATGGTGGTGGTGGGGATG 410

Qy     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187
           |  |      ||||   |||| |||| | | |       |    || || || || |   |
Db     411 GTGATGATGGTGGTGGTGGTAATGGTGGTGGTAATGATGGTGGTGATGATGGTGGTGATG 470

Qy     188 GTGGATTAATTGGT 201
           ||||    | ||||
Db     471 GTGGTGGTAATGGT 484


RESULT 7
US-09-815-242-6926
; Sequence 6926, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6926
;   LENGTH: 1671
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1671)
US-09-815-242-6926

  Query Match             7.9%;  Score 38.4;  DB 10;  Length 1671;
  Best Local Similarity   55.1%;  Pred. No. 0.063;
  Matches   75;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      18 GCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAG 77
           ||||||| | | | |||   ||   |       |  ||||||   || | |  |  |||
Db     396 GCAAGGTTGGAATGTAAAAGTTGAAACCCGTGGTCAAGTTGGGGCTGGCAATGAAATTAC 455

Qy      78 TCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTT 137
           |   ||||||||| ||| ||   |  ||   ||||||  |  | |  || |  ||  ||
```

```
Db     456 TGTGGAAGAAGTCGCAGCGGCTGATTTGGTATTTGTTGCAGCAGATATTGACGTGCCTTT 515

Qy     138 TGGTAAAGGTAGTGGT 153
           | ||||  ||  |||
Db     516 AGATAAATTTAAAGGT 531


RESULT 8
US-09-864-761-20699
; Sequence 20699, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20699
;   LENGTH: 446
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007249.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  13
```

```
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
;   OTHER INFORMATION: NT HIT: AL161539.2, EVALUE 3.70e+00
US-09-864-761-20699

  Query Match             7.9%;  Score 38.2;  DB 10;  Length 446;
  Best Local Similarity   54.7%;  Pred. No. 0.036;
  Matches   76;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy      75 TAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCT 134
           | || || | |  |  || | ||    ||| | || || ||| |||| || |||
Db     170 TGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGA 229

Qy     135 GTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATT 194
             ||||   ||| ||||| | | ||     |  | ||||||||| || |   |||||
Db     230 TGGTGGTGGTGGTGGTGGTGGTGGTGATGATGGTAGTGGTGGTGATGGTGGTGGTGGTGG 289

Qy     195 AATTGGTTCTAAAATCGGT 213
           ||||  | |    |||
Db     290 TGGTGGTGGTGATGGTGGT 308


RESULT 9
US-09-864-761-19241
; Sequence 19241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19241
;  LENGTH: 1075
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AL078472.1
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  27
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  19
;  OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  34
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  28
;  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  43
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;  OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  25
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  18
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  21
;  OTHER INFORMATION: EST_HUMAN HIT: AV739739.1, EVALUE 1.00e+00
;  OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-19
US-09-864-761-19241

  Query Match             7.8%;  Score 37.8;  DB 10;  Length 1075;
  Best Local Similarity   51.5%;  Pred. No. 0.076;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      22 GGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGT 81
           ||| || ||   | |  |    |||    |   | ||   |||   |       | || ||
Db     181 GGTGGTGGTGGTAGTGGTGATGGTGGTGGTGGTGGTGATGGTGGTGATGGTGGTGGTGGT 240

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
             | |  |  || | ||    ||| | || || ||| |||| || |||  |      ||||
Db     241 GATGGTGGTGGTGGTGGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGATGGTAGTGGT 300

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTG 190
           | ||| ||||| | | ||     |    ||||| || || |   ||||
Db     301 GATGGTGGTGGTGGTGGTGATGGTGGTGATGGTGATGGTGGTGGTGGTG 349


RESULT 10
US-09-864-761-2513
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2513
;   LENGTH: 1403
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  21
US-09-864-761-2513

  Query Match             7.8%;  Score 37.8;  DB 10;  Length 1403;
  Best Local Similarity   51.5%;  Pred. No. 0.088;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      22 GGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGT 81
           ||| || ||   | |  |    |||    |   | ||   |||   |       | || ||
Db     531 GGTGGTGGTGGTAGTGGTGATGGTGGTGGTGGTGGTGATGGTGGTGATGGTGGTGGTGGT 590

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
             |  |  || | ||    ||| | || || ||| |||| || |||  |      ||||
Db     591 GATGGTGGTGGTGGTGGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGATGGTAGTGGT 650

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTG 190
           | ||| ||||| | | ||     |    ||||| || || |   ||||
Db     651 GATGGTGGTGGTGGTGGTGATGGTGGTGATGGTGATGGTGGTGGTGGTG 699


RESULT 11
US-09-864-761-3121
; Sequence 3121, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3121
;   LENGTH: 470
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL049749.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.4
US-09-864-761-3121

  Query Match             7.7%;  Score 37.6;  DB 10;  Length 470;
  Best Local Similarity   57.8%;  Pred. No. 0.057;
  Matches   67;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      76 AGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTG 135
           |||  ||| ||||  || | ||    ||| | || || ||| | || || |||  |
Db      86 AGTGATCATGAAGGTGGTGGTGATGGTGGTGATGGTGGTGGTGTTGATGATGGTGATGGT 145

Qy     136 TTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGG 191
           ||||  | | | ||||| |   ||     | || || ||||| || |   | |||
Db     146 GGTGGTGATGATGGTGGTGGTAGTGGTGGTGACAGTGATGGTGGTGGTAGTGATGG 201


RESULT 12
US-09-864-761-16846
; Sequence 16846, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
```

```
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16846
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007263.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.8
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1
;   OTHER INFORMATION: NT HIT: AF164623.1, EVALUE 5.10e-01
;   OTHER INFORMATION: EST_HUMAN HIT: AW846120.1, EVALUE 1.00e-02
;   OTHER INFORMATION: SWISSPROT HIT: Q12329, EVALUE 4.80e+00
US-09-864-761-16846

  Query Match             7.7%;  Score 37.4;  DB 10;  Length 249;
  Best Local Similarity   55.9%;  Pred. No. 0.046;
  Matches   71;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      75 TAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCT 134
           | || ||   |  |  || | ||    ||| | || || ||| |||| || ||| |
Db      10 TGGTGGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTCGTGG 69

Qy     135 GTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATT 194
             ||||  |||  |||| | ||||   | |   |||||| || || |   |||||
Db      70 TGGTGGTGTTGGTGTTGGTGGTGTTGGTGTTGGTGTTGGTGATGGTGGTGGTGGTGGTGG 129

Qy     195 AATTGGT 201
              ||||
Db     130 TCGTGGT 136


RESULT 13
US-09-864-761-6
; Sequence 6, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6
;   LENGTH: 476
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007263.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.8
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.6
```

```
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1
US-09-864-761-6

  Query Match             7.7%;  Score 37.4;  DB 10;  Length 476;
  Best Local Similarity   52.2%;  Pred. No. 0.066;
  Matches   83;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy      55 GTTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTT 114
           | |||  ||| |        |||| ||   |  |  || | ||    ||| | || || |
Db     260 GGTGGTGGTGGCGGTGGTGGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 319

Qy     115 GGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGT 174
           || |||| |  |||        |||||   ||| ||| | | ||||   | |    ||||
Db     320 GGTGGTGGTCGTGGTGGTGGTGTTGGTGTTGGTGGTGTTGGTGTTGGTGTTGGTGATGGT 379

Qy     175 GGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGT 213
           |||| || |   ||| |      ||||  | |    |||
Db     380 GGTGGTGGTGGTGGTCGTGGTGGTGGTAGTGATGGTGGT 418


RESULT 14
US-09-864-761-20241/c
; Sequence 20241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20241
;  LENGTH: 1635
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AC006547.9
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
;  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
;  OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
;  OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-20241

  Query Match             7.6%;  Score 36.8;  DB 10;  Length 1635;
  Best Local Similarity   49.5%;  Pred. No. 0.2;
  Matches   95;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      22 GGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGT 81
           ||| |||||     |  |    |||    |   | |||  |||       |   | || ||
Db    1009 GGTGGTAGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTGGTAGT 950

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
             | |  |  || | ||    |||   || || ||| |||| || |||         ||||
Db     949 GATGGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTGATGGT 890

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
             ||| ||||| | | ||      |    |||||||  || |   |||||      ||||
Db     889 GGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGT 830

Qy     202 TCTAAAATCGGT 213
             ||     |||
Db     829 GGTAGTGATGGT 818


RESULT 15
US-09-864-761-3471/c
; Sequence 3471, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3471
;   LENGTH: 1973
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006547.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  17
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
US-09-864-761-3471

  Query Match             7.6%;  Score 36.8;  DB 10;  Length 1973;
  Best Local Similarity   49.5%;  Pred. No. 0.22;
  Matches   95;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy      22 GGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCCAGAACTTTAGTCGT 81
           ||| || ||   | |  |    |||    |   | |||| |||         |||| ||
Db     824 GGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGCAGTGATGGTGGTGGTAGTGGT 765

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
             || |  || | ||    | | | || || ||| |||| || |||         ||||
Db     764 AGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGATGATGGT 705

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           ||| ||||| |   ||     |    |||||||| || |   | |||       ||||
Db     704 GGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTGATGGTGATGGTGGT 645

Qy     202 TCTAAAATCGGT 213
             | |    |||
Db     644 AGTGATGGTGGT 633


Search completed: February 22, 2003, 18:59:58
Job time : 103.285 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:08:35 ; Search time 135.699 Seconds
(without alignments)
8065.404 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagaggatgtttgca.....aagtcatttcaacagaaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	486	100.0	486	22	AAF86246	OspA 17kD antigen
	2	486	100.0	486	24	ABK52401	DNA encoding outer
	3	486	100.0	489	22	AAH79040	Piscirickettsia sa
	4	295	60.7	483	22	AAF86247	DNA sequence of E.
	5	295	60.7	483	24	ABK52402	DNA encoding Esche
	6	295	60.7	768	22	AAF86248	DNA sequence of cl
	7	295	60.7	768	24	ABK52403	E. coli codon opti
	8	73.2	15.1	110	22	AAF86252	PCR primer #2 used
	9	73.2	15.1	110	24	ABK52407	Escherichia coli c
c	10	71.6	14.7	118	22	AAF86254	PCR primer #4 used
c	11	71.6	14.7	118	24	ABK52409	Escherichia coli c
c	12	69	14.2	102	22	AAF86255	PCR primer #5 used
c	13	69	14.2	102	24	ABK52410	Escherichia coli c
	14	62	12.8	94	22	AAF86253	PCR primer #3 used
	15	62	12.8	94	24	ABK52408	Escherichia coli c
	16	56.2	11.6	454	20	AAZ06832	Rickettsia 17 kD o
c	17	46.8	9.6	110	22	AAF86256	PCR primer #6 used
c	18	46.8	9.6	110	24	ABK52411	Escherichia coli c
c	19	44	9.1	309	23	ABL13299	Drosophila melanog
	20	44	9.1	2309	23	ABL13298	Drosophila melanog
	21	43.2	8.9	155074	24	ABN85735	Human genomic regi
	22	42.6	8.8	790	11	AAQ03870	H.influenzae DNA f
	23	42.6	8.8	850	9	AAN80227	Sequence of Haemop
c	24	41.2	8.5	107602	24	AAK99657	DNA of the PAC clo
c	25	41.2	8.5	107612	24	ABL54503	Human PAC clone se
c	26	40.4	8.3	1509	23	ABL20279	Drosophila melanog
	27	40.4	8.3	5162	23	ABL20278	Drosophila melanog
	28	39.8	8.2	897	24	ABQ68544	Listeria monocytog
c	29	39.8	8.2	2226	23	AAS78586	DNA encoding novel
c	30	39.8	8.2	2292	21	AAC45883	Arabidopsis thalia
	31	39.8	8.2	7409	24	ABQ71057	Listeria monocytog
	32	39.6	8.1	522	22	ABA49593	Human breast cell
	33	39.6	8.1	522	22	ABA67501	Human foetal liver
	34	39.6	8.1	522	22	ABA34580	Probe #13046 for g
	35	39.6	8.1	522	22	AAK15928	Human brain expres
	36	39.6	8.1	522	22	AAK41670	Human bone marrow
	37	39.6	8.1	522	22	AAI22424	Probe #12357 for g
	38	39.6	8.1	522	22	AAI47715	Probe #16401 used
	39	39.6	8.1	522	22	AAI08110	Probe #8101 used t
	40	39.6	8.1	522	24	ABS15679	Human genome-deriv
c	41	39.4	8.1	1395	23	AAS89524	DNA encoding novel
	42	39	8.0	350	22	ABA72229	Human foetal liver
	43	39	8.0	350	22	ABA38104	Probe #16570 for g
	44	39	8.0	350	22	AAK20651	Human brain expres
	45	39	8.0	350	22	AAK46795	Human bone marrow

ALIGNMENTS

RESULT 1
AAF86246
ID AAF86246 standard; DNA; 486 BP.
XX
AC AAF86246;
XX
DT 11-JUL-2001 (first entry)
XX
DE OspA 17kD antigen gene.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; ds.
XX
OS Piscirickettsia salmonis.
XX
FH Key Location/Qualifiers
FT CDS 1..486
FT /*tag= a
FT /partial
FT /product= "OspA"
FT /note= "Genus specific 17kDa antigen, the sequence does
FT not include a stop codon"
XX
PN CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX

```
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81126.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Disclosure; Fig 2B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents DNA encoding the P. salmonis OspA protein. An OspA
CC   protein with an N-terminal fusion partner is used in a vaccine to create
CC   an anti-OspA antibody response.
XX
SQ   Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

  Query Match             100.0%;  Score 486;  DB 22;  Length 486;
  Best Local Similarity   100.0%;  Pred. No. 4e-137;
  Matches  486;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60

Qy      61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120

Qy     121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180

Qy     181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240

Qy     241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300

Qy     301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360

Qy     361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420

Qy     421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480

Qy     481 GAAAAA 486
          ||||||
Db     481 GAAAAA 486


RESULT 2
ABK52401
ID   ABK52401 standard; DNA; 486 BP.
XX
AC   ABK52401;
XX
DT   12-AUG-2002  (first entry)
XX
DE   DNA encoding outer surface lipoprotein OspA.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag=  a
FT                   /product= "OspA"
FT                   /note= "Outer surface lipoprotein"
FT                   /partial= "No stop codon given"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97867.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 19; Fig 2; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Piscirickettsia salmonis outer surface
CC   lipoprotein, OspA, used in the creation of the vaccine described in the
CC   invention.
XX
SQ   Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

  Query Match             100.0%;  Score 486;  DB 24;  Length 486;
  Best Local Similarity   100.0%;  Pred. No. 4e-137;
  Matches  486;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60

Qy      61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120

Qy     121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180
```

```
Db    121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180

Qy    181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240
Db    181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240

Qy    241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300
Db    241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300

Qy    301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360
Db    301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360

Qy    361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420
Db    361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420

Qy    421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480
Db    421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480

Qy    481 GAAAAA 486
Db    481 GAAAAA 486
```

```
RESULT 3
AAH79040
ID   AAH79040 standard; cDNA; 489 BP.
XX
AC   AAH79040;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Piscirickettsia salmonis polynucleotide P10.6.
XX
KW   Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW   septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW   ATCC VR-1361; ss.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..489
FT                   /*tag=  a
XX
PN   WO200168865-A2.
XX
PD   20-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-GB01055.
XX
PR   11-MAR-2000; 2000GB-0005838.
PR   01-JUL-2000; 2000GB-0016080.
PR   01-JUL-2000; 2000GB-0016082.
PR   29-JUL-2000; 2000GB-0018599.
XX
PA   (AQUA-) AQUA HEALTH EURO LTD.
XX
PI   Simard N,  Brouwers H,  Jones S,  Griffiths S,  Valenzuela P;
PI   Burzio L;
XX
DR   WPI; 2001-639050/73.
DR   P-PSDB; AAG78025.
XX
PT   New nucleic acids encoding an amino acid sequence homologous to the
PT   surface antigen present on Piscirickettsia salmonis are useful to
PT   protect fish against piscirickettsiosis  -
XX
PS   Claim 3; Fig 4; 25pp; English.
XX
CC   The invention relates to nucleic acid sequences and the encoded protein
CC   of a least part of the surface antigen present on Piscirickettsia
CC   salmonis for production of a vaccine with antibacterial activity to
CC   protect fish against P. salmonis which causes piscirickettsiosis, also
CC   known as salmonid rickettsial septicaemia.
XX
SQ   Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other;
```

```
  Query Match           100.0%;  Score 486;  DB 22;  Length 489;
  Best Local Similarity  100.0%;  Pred. No. 4e-137;
  Matches  486;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60
Db      1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60

Qy     61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120
Db     61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120

Qy    121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180
Db    121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180

Qy    181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240
Db    181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240

Qy    241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300
Db    241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300

Qy    301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360
Db    301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360

Qy    361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420
Db    361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420

Qy    421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480
Db    421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480

Qy    481 GAAAAA 486
Db    481 GAAAAA 486
```

```
RESULT 4
AAF86247
ID   AAF86247 standard; DNA; 483 BP.
XX
AC   AAF86247;
XX
DT   11-JUL-2001  (first entry)
XX
DE   DNA sequence of E. coli optimised ospA gene 17E2.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; ds.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..483
FT                   /*tag=  a
FT                   /partial
FT                   /product= "OspA"
FT                   /note= "Genus specific 17kDa antigen, the sequence does
FT                         not include a stop codon"
```

```
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81127.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4C; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents DNA which has been optimised for expression in
CC   Escherichia coli to encode the P. salmonis OspA protein. An OspA protein
CC   with an N-terminal fusion partner is used in a vaccine to create an
CC   anti-OspA antibody response.
XX
SQ   Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

  Query Match             60.7%;  Score 295;  DB 22;  Length 483;
  Best Local Similarity   76.0%;  Pred. No. 3e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67
Db       5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64

Qy      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127
Db      65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124

Qy     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187
Db     125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184

Qy     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247
Db     185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244

Qy     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307
Db     245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304

Qy     308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367
Db     305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364

Qy     368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427
Db     365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424

Qy     428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486
```

```
Db     425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483


RESULT 5
ABK52402
ID   ABK52402 standard; DNA; 483 BP.
XX
AC   ABK52402;
XX
DT   12-AUG-2002  (first entry)
XX
DE   DNA encoding Escherichia coli codon optimised OspA, 17e2.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds; 17e2.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..483
FT                   /*tag=  a
FT                   /product= "17e2"
FT                   /note= "Escherichia coli codon optimised OspA"
FT                   /partial
FT                   /note= "No stop codon given"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97868.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 16; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Escherichia coli codon optimised outer surface
CC   lipoprotein OspA (17e2) used in the creation of the vaccine described in
CC   the invention.
XX
SQ   Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

  Query Match             60.7%;  Score 295;  DB 24;  Length 483;
  Best Local Similarity   76.0%;  Pred. No. 3e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67
```

```
Db      5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64

Qy     68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Db     65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124

Qy    128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Db    125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184

Qy    188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Db    185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244

Qy    248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Db    245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304

Qy    308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Db    305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364

Qy    368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Db    365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424

Qy    428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486

Db    425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483


RESULT 6
AAF86248
ID   AAF86248 standard; DNA; 768 BP.
XX
AC   AAF86248;
XX
DT   11-JUL-2001  (first entry)
XX
DE   DNA sequence of c17e2 ospA construct with N-terminal fusion partner.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; ds.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..768
FT                   /*tag=  a
FT                   /partial
FT                   /product= "C17E2 OspA with N-terminal fusion partner"
FT                   /note= "No stop codon is given"
FT   misc_feature    1..285
FT                   /*tag=  b
FT                   /note= "DNA encoding undefined N-terminal fusion partner"
FT   misc_feature    286..768
FT                   /*tag=  c
FT                   /note= "Optimised OspA construct c17E2"
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81128.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 4; Fig 5; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents P. salmonis OspA DNA termed C17E2 optimised for
CC   expression in Escherichia coli fused to DNA encoding an undefined
CC   N-terminal fusion partner. The protein encoded by this fusion construct
CC   is used in a vaccine to create an anti-OspA antibody response.
XX
SQ   Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

  Query Match            60.7%;  Score 295;  DB 22;  Length 768;
  Best Local Similarity  76.0%;  Pred. No. 3.7e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy      8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Db    290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349

Qy     68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Db    350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409

Qy    128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Db    410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469

Qy    188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Db    470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529

Qy    248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Db    530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589

Qy    308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Db    590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649

Qy    368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Db    650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709

Qy    428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486

Db    710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768


RESULT 7
ABK52403
ID   ABK52403 standard; DNA; 768 BP.
XX
AC   ABK52403;
XX
DT   12-AUG-2002  (first entry)
XX
DE   E. coli codon optimised OspA, 17e2 with N-terminal fusion DNA.
```

```
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds; 17e2.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..768
FT                   /*tag= a
FT                   /product= "17e2"
FT                   /note= "Escherichia coli codon optimised OspA with
FT                        N-terminal fusion"
FT                   /partial
FT                   /note= "No stop codon given"
FT   sig_peptide     1..285
FT                   /*tag= a
FT                   /note= "This sequence represents the N-terminal fusion"
FT   mat_peptide     286..768
FT                   /*tag= c
FT                   /label= "mature_17e2"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97869.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 16; Fig 5; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Escherichia coli codon optimised outer surface
CC   lipoprotein OspA (17e2) with an N-terminal fusion used in the creation of
CC   the vaccine described in the invention.
XX
SQ   Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

  Query Match             60.7%;  Score 295;  DB 24;  Length 768;
  Best Local Similarity   76.0%;  Pred. No. 3.7e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67
           | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||
Db     290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349

Qy      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127
           ||||||| || || || ||||| || || || || || || ||||| ||||||||||| |

Db     350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409

Qy     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187
           |||||||||| ||||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db     410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469

Qy     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247
           | ||  | ||||| |||||||||||||||     ||||| |||||||||||||| || || |
Db     470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529

Qy     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307
           |||||  | ||||||| || ||||| || || || || ||||||||||||||| || || |
Db     530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589

Qy     308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367
           | || || ||   ||| || || || || || ||||||||||||||| || || || || |
Db     590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649

Qy     368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427
           | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db     650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709

Qy     428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486
           | |||||||| || ||||  || |||||||| || ||||| || |||   || ||||||
Db     710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768


RESULT 8
AAF86252
ID   AAF86252 standard; DNA; 110 BP.
XX
AC   AAF86252;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #2 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
```

```
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

  Query Match             15.1%;  Score 73.2;  DB 22;  Length 110;
  Best Local Similarity   79.1%;  Pred. No. 2.4e-12;
  Matches   87;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      56 TTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTG 115
           | || || |||||||||||| || || || ||||| || || || || || || ||||| |
Db       1 TGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGG 60

Qy     116 GCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATG 165
           |||||||||| |||||||||| ||||||||   |||||| || || |||
Db      61 GCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGGCGATG 110


RESULT 9
ABK52407
ID   ABK52407 standard; DNA; 110 BP.
XX
AC   ABK52407;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #2.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Example 3; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence represents an oligonucleotide used to construct an
CC   Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC   used in the creation of the vaccine described in the invention.
XX
SQ   Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

  Query Match             15.1%;  Score 73.2;  DB 24;  Length 110;
  Best Local Similarity   79.1%;  Pred. No. 2.4e-12;
  Matches   87;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      56 TTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTG 115
           | || || |||||||||||| || || || ||||| || || || || || || ||||| |
Db       1 TGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGG 60

Qy     116 GCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATG 165
           |||||||||| |||||||||| ||||||||   |||||| || || |||
Db      61 GCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGGCGATG 110


RESULT 10
AAF86254/c
ID   AAF86254 standard; DNA; 118 BP.
XX
AC   AAF86254;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #4 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;

  Query Match             14.7%;  Score 71.6;  DB 22;  Length 118;
  Best Local Similarity   75.4%;  Pred. No. 7.7e-12;
  Matches   89;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy     211 GGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAAAAGGTAAAA 270
           |||||    ||||| |||||||||||||| || || ||||||  | |||||||| || |||
Db     118 GGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAA 59
```

```
Qy    271 GCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGTGTTGAGC 328
          || || || || || |||||||||||||| || || || || || ||   ||| || |
Db     58 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAAC 1
```

RESULT 11
ABK52409/c
ID ABK52409 standard; DNA; 118 BP.
XX
AC ABK52409;
XX
DT 12-AUG-2002 (first entry)
XX
DE Escherichia coli codon optimised OspA 17e2, oligonucleotide #4.
XX
KW Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW Rickettisial disease; 17e2; ss.
XX
OS Synthetic.
XX
PN CA2339327-A1.
XX
PD 15-MAR-2002.
XX
PF 19-MAR-2001; 2001CA-2339327.
XX
PR 15-SEP-2000; 2000US-0677374.
XX
PA (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZY/) KUZYK M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
XX
DR WPI; 2002-455221/49.
XX
PT Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT administration of an outer surface lipoprotein (OspA) of a bacterial
PT strain, as a vaccine -
XX
PS Example 3; Fig 4; 55pp; English.
XX
CC The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen Piscirickettsia Salmonis
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the OspA (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (OspA) of
CC bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicaemia and other related
CC Rickettisial diseases caused by either a virus, bacteria or parasite.
CC This sequence represents an oligonucleotide used to construct an
CC Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC used in the creation of the vaccine described in the invention.
XX
SQ Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;

```
  Query Match             14.7%;  Score 71.6;  DB 24;  Length 118;
  Best Local Similarity   75.4%;  Pred. No. 7.7e-12;
  Matches   89;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    211 GGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAAAAGGTAAAA 270
          |||||    ||||| |||||||||||||| || || ||||||  | ||||||| || |||
Db    118 GGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAA 59

Qy    271 GCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGTGTTGAGC 328
          || || || || || |||||||||||||| || || || || || ||   ||| || |
Db     58 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAAC 1
```

RESULT 12
AAF86255/c
ID AAF86255 standard; DNA; 102 BP.
XX
AC AAF86255;
XX
DT 11-JUL-2001 (first entry)
XX
DE PCR primer #5 used in cloning an optimisation of OspA gene.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS Piscirickettsia salmonis.
XX
PN CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZY/) KUZYK M A.
XX
PI Kay WW, Burian J, Kuzyk MA;
XX
DR WPI; 2001-316844/34.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OspA protein of Piscirickettsia salmonis -
XX
PS Example 3; Fig 4B; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis OspA gene. The OspA gene is used in the method of the
CC invention.
XX
SQ Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

```
  Query Match             14.2%;  Score 69;  DB 22;  Length 102;
  Best Local Similarity   80.2%;  Pred. No. 4.4e-11;
  Matches   81;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy    337 ACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAA 396
          || |||||||||||||| || || || || || ||||| || || || ||||||||||
Db    102 ACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAG 43

Qy    397 AAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCAC 437
          || || ||||| ||||| ||||| || || || ||||||||
Db     42 AAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCAC 2
```

RESULT 13
ABK52410/c
ID ABK52410 standard; DNA; 102 BP.
XX
AC ABK52410;
XX
DT 12-AUG-2002 (first entry)

```
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #5.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Example 3; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence represents an oligonucleotide used to construct an
CC   Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC   used in the creation of the vaccine described in the invention.
XX
SQ   Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

  Query Match             14.2%;  Score 69;  DB 24;  Length 102;
  Best Local Similarity   80.2%;  Pred. No. 4.4e-11;
  Matches   81;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy    337 ACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAA 396
          || |||||||||||||| || || || || || ||||| || || || ||||||||||
Db    102 ACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAG 43

Qy    397 AAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCAC 437
          || || ||||| ||||| ||||| || || || ||||||||
Db     42 AAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCAC 2


RESULT 14
AAF86253
ID   AAF86253 standard; DNA; 94 BP.
XX
AC   AAF86253;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #3 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

  Query Match             12.8%;  Score 62;  DB 22;  Length 94;
  Best Local Similarity   78.7%;  Pred. No. 5.6e-09;
  Matches   74;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy    142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
          |||||   |||||| || || |||||||| || ||||| ||| |||| ||  | |||||
Db      1 AAAGGCTCTGGTCGTGTGGCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 60

Qy    202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATA 235
          ||||||||||||||    ||||| ||||||||||
Db     61 TCTAAAATCGGTCAGAGCATGGACCAGCAGGATA 94


RESULT 15
ABK52408
ID   ABK52408 standard; DNA; 94 BP.
XX
AC   ABK52408;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #3.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
```

PF 19-MAR-2001; 2001CA-2339327.
XX
PR 15-SEP-2000; 2000US-0677374.
XX
PA (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZY/) KUZYK M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
XX
DR WPI; 2002-455221/49.
XX
PT Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT administration of an outer surface lipoprotein (OspA) of a bacterial
PT strain, as a vaccine -
XX
PS Example 3; Fig 4; 55pp; English.
XX
CC The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen Piscirickettsia Salmonis
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the OspA (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (OspA) of
CC bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicaemia and other related
CC Rickettisial diseases caused by either a virus, bacteria or parasite.
CC This sequence represents an oligonucleotide used to construct an
CC Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC used in the creation of the vaccine described in the invention.
XX
SQ Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

Query Match 12.8%; Score 62; DB 24; Length 94;
Best Local Similarity 78.7%; Pred. No. 5.6e-09;
Matches 74; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy    142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
          |||||   |||||| || || |||||||| || ||||| ||| |||| ||  | |||||
Db      1 AAAGGCTCTGGTCGTGTGGCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 60

Qy    202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATA 235
          |||||||||||||    ||||| ||||||||||
Db     61 TCTAAAATCGGTCAGAGCATGGACCAGCAGGATA 94
```

Search completed: February 22, 2003, 14:44:50
Job time : 141.699 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:11:08 ; Search time 1355.32 Seconds
(without alignments)
10435.915 Million cell updates/sec

Title: US-09-677-374-1
Perfect score: 486
Sequence: 1 atgaacagaggatgtttgca.....aagtcatttcaacagaaaaa 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	486	100.0	489	6	AX252413	AX252413 Sequence
2	486	100.0	4983	1	AF184152	AF184152 Piscirick
3	79	16.3	448	1	RSU76907	U76907 Rickettsia
4	70.6	14.5	237523	1	RPXX04	AJ235273 Rickettsi
5	68.4	14.1	416	1	AF031534	AF031534 Rickettsi
6	65.8	13.5	535	1	RIRANT17KB	M28481 R.typhi 17K
7	65.8	13.5	537	1	RIR17KGCA	D16515 Rickettsia
8	64.6	13.3	539	1	RIRANT17KA	M28480 R.conorii 1
9	64.6	13.3	10127	1	AE008675	AE008675 Rickettsi
10	64.2	13.2	539	1	RIRANT17KD	M28479 R.rickettsi
11	64.2	13.2	620	1	RIRANTRR	M16486 R.rickettsi
12	62.2	12.8	434	1	MRI269516	AJ269516 Male-kill
13	62.2	12.8	434	1	MRI269517	AJ269517 Male-kill
14	62.2	12.8	434	1	MRI269518	AJ269518 Male-kill
15	59.4	12.2	546	1	AF195118	AF195118 Rickettsi
16	59	12.1	489	1	RRU11020	U11020 Rickettsia
17	58.2	12.0	415	1	AF027124	AF027124 Rickettsi
18	57.8	11.9	532	1	RIRANT17KC	M28482 R.prowazeki
19	57.6	11.9	411	1	AF060704	AF060704 Rickettsi
20	57.6	11.9	411	1	AF060706	AF060706 Rickettsi
21	57.4	11.8	489	1	RPU17008	U17008 Rickettsia
22	57.4	11.8	498	1	RAU11013	U11013 Rickettsia
23	56.2	11.6	454	1	AF181036	AF181036 Rickettsi
24	55.6	11.4	489	1	RMU11017	U11017 Rickettsia
25	54.4	11.2	395	1	U04162	U04162 Rickettsia
26	54.2	11.2	492	1	RIRTRAPRO	M74042 R.australis
27	52.6	10.8	491	1	RIRGENSEQ	M99391 Rickettsia
c 28	49	10.1	230993	2	AC114815	AC114815 Homo sapi
29	47.8	9.8	394	1	AF260571	AF260571 Rickettsi
c 30	47.8	9.8	7218	6	I66494	I66494 Sequence 14
c 31	46.6	9.6	206114	2	AC118508	AC118508 Rattus no
c 32	46.4	9.5	111824	9	AL160412	AL160412 Human DNA
33	46.4	9.5	206114	2	AC118508	AC118508 Rattus no
c 34	46.2	9.5	165087	2	AC102004	AC102004 Mus muscu
c 35	45.4	9.3	68651	9	AC027343	AC027343 Homo sapi
c 36	45.4	9.3	121829	2	AC103214	AC103214 Rattus no
c 37	45.4	9.3	152198	2	AC128626	AC128626 Rattus no
c 38	45.2	9.3	162165	2	AC127626	AC127626 Rattus no
c 39	45	9.3	61052	2	AC117074	AC117074 Dictyoste
c 40	45	9.3	61052	2	AC123513	AC123513 Dictyoste
c 41	45	9.3	156370	2	AL158162	AL158162 Homo sapi
42	45	9.3	166464	9	AL158212	AL158212 Human DNA
c 43	44.8	9.2	212383	2	AC119030	AC119030 Rattus no
c 44	44.8	9.2	217416	2	AC103538	AC103538 Rattus no
45	44.6	9.2	13123	1	AE003872	AE003872 Xylella f

ALIGNMENTS

RESULT 1
AX252413
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168865.
ACCESSION AX252413
VERSION AX252413.1 GI:15985721
KEYWORDS .
SOURCE Piscirickettsia salmonis.
ORGANISM Piscirickettsia salmonis
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
Piscirickettsia.
REFERENCE 1 (bases 1 to 489)
AUTHORS Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and
Burzio,L.
TITLE Fish vaccine against piscirickettsia salmonis

```
  JOURNAL   Patent: WO 0168865-A 5 20-SEP-2001;
            Aqua Health (Europe) Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Piscirickettsia salmonis"
                     /db_xref="taxon:1238"
BASE COUNT      139 a     79 c    144 g    127 t
ORIGIN

  Query Match           100.0%;  Score 486;  DB 6;  Length 489;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-129;
  Matches  486;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60

Qy      61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120

Qy     121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180

Qy     181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240

Qy     241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300

Qy     301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360

Qy     361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420

Qy     421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480

Qy     481 GAAAAA 486
           ||||||
Db     481 GAAAAA 486


RESULT 2
AF184152
LOCUS       AF184152               4983 bp    DNA     linear   BCT 01-OCT-2000
DEFINITION  Piscirickettsia salmonis alanine racemase (alr) gene, partial cds;
            and BAX (bax), DNA repair enzyme RadA (radA), 17 kDa antigen
            (ospA), and transposase (tnpA) genes, complete cds.
ACCESSION   AF184152
VERSION     AF184152.1  GI:10441343
KEYWORDS    .
SOURCE      Piscirickettsia salmonis.
  ORGANISM  Piscirickettsia salmonis
            Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
            Piscirickettsia.
REFERENCE   1  (bases 1 to 4983)
  AUTHORS   Kuzyk,M.A., Burian,J., Thornton,J.C. and Kay,W.W.
  TITLE     Identification of a genus-common Rickettsial surface antigen in the
            salmonid pathogen Piscirickettsia salmonis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4983)
  AUTHORS   Kuzyk,M.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Biochemistry & Microbiology, University of
            Victoria, Petch Building, PO Box 3055, Victoria, BC V8W 3P6, Canada

FEATURES             Location/Qualifiers
     source          1. .4983
                     /organism="Piscirickettsia salmonis"
                     /strain="LF-89"
                     /db_xref="taxon:1238"
                     /clone="pB12"
     source          3560. .4849
                     /organism="Piscirickettsia salmonis"
                     /strain="LF-89"
                     /db_xref="taxon:1238"
                     /clone="pB12"
                     /insertion_seq="insertion sequence-like element"
     gene            <1. .501
                     /gene="alr"
     CDS             <1. .501
                     /gene="alr"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="alanine racemase"
                     /protein_id="AAG17001.1"
                     /db_xref="GI:10441347"
                     /translation="NSAAIFNFSYERYDWVRPGIMLYGISPFADKNGVDLELQPVMHV
                     VSRLISVKQLRQGESVGYGATWQCPEDMQVGILSLGYGDGYPRLAASGTPFLVRGQRC
                     ALIGRVSMDMIAIDLRRCPDAGVGEAVTVWGQDLPVEEIARHVGTIAYELVCNMPLRA
                     PYIWQE"
     gene            complement(526. .1257)
                     /gene="bax"
     CDS             complement(526. .1257)
                     /gene="bax"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="BAX"
                     /protein_id="AAG17002.1"
                     /db_xref="GI:10441348"
                     /translation="MLSLAVITTAYLAQQPLFSAPTSKTQINTAYSQDIQSIKQEFII
                     KITQSENTINQIILQKRQHILSLIQKKSLSKQEITQLQALAKQYKLNNLHFKHPQDFS
                     LLLKRVDIIPNALISAQAINESNWGRSRFAVEGNNFFGMRCHYPGCGIIPKARPANNH
                     WEVANYSSMTASVRAYIHTLNTHNAYQALRDLRAHMRANHQDVSAFKLAEGLTAYSIK
                     GTKYVELIQTIITRYLSKESTSQIP"
     gene            1373. .2743
                     /gene="radA"
                     /evidence=experimental
     CDS             1373. .2743
                     /gene="radA"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA repair enzyme RadA"
                     /protein_id="AAG16998.1"
                     /db_xref="GI:10441344"
                     /translation="MSKQKNQYVCSDCGGIAKKWLGQCPHCQQWNSLSEVKEVLPNRP
                     GRSQRFAGFAGIDAPKVKCLSEITPEQISRQPVGICEFDRVLGGGLVHGAVILLGGDP
                     GIGKSTLLLQTSVNCTQFGKVLYVTGEESLEQVTLRSKRLGLSQDVDLRLLAETQVER
                     ILKAAEIEQPKVLIVDSIQTIFTESLQSAPGGVAQVRESAAILTQFAKRTGTCLFLVG
                     HVTKEGALAGPRVLEHMVDTVLYFEGEQDSRFRLLRAVKNRFGAANELGIFAMTETGL
                     KTVSNPSAIFLSRYEDLQPGSVVMVAWEGTRPLLVEVQALVDESHSSNPRRIAVGLDQ
                     QRLAMLLAVLNRHGGIASYDQDVFINVVGGMKITETAADLALLLACVSSLRGKALSQE
                     LIVFGEIGLSGEIRPVQRGLERIKEAVKHGFTQAIVPLANCPKQKVDGIEIVGVKHLE
                     DALATL"
     gene            2834. .3322
                     /gene="ospA"
                     /evidence=experimental
     CDS             2834. .3322
                     /gene="ospA"
                     /note="similar to the rickettsial 17 kDa antigen"
                     /codon_start=1
                     /transl_table=11
                     /product="17 kDa antigen"
                     /protein_id="AAG17000.1"
                     /db_xref="GI:10441346"
                     /translation="MNRGCLQGSSLIIISVFLVGCAQNFSRQEVGAATGAVVGGVAGQ
                     LFGKGSGRVAMAIGGAVLGGLIGSKIGQSMDQQDKIKLNQSLEKVKAGQVTRWRNPDT
```

```
                     GNSYSVEPVRTYQRYNKQERRQQYCREFQQKAMIAGQKQEIYGTACRQPDGRWQVIST
                     EK"
     repeat_region   3560. .3846
     gene            complement(3842. .4561)
                     /gene="tnpA"
     CDS             complement(3842. .4561)
                     /gene="tnpA"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAG16999.1"
                     /db_xref="GI:10441345"
                     /translation="MDLTLISLFCVIDDFCQELLPQWNAILLEDTNKKRNKPSQMSTS
                     EIMTIMIYFHKSNYRNFKMYYLHVIKGSMVKYFPNSVSYNRFVELMPSILLPLCFFIA
                     AQGKTATGIYFVDSTILRVCHEKRASQNRAFKGLAKKSKSTMGWYYGFKLHIIVNDMG
                     ELMAFKMSKATTDDRVVLPKMAENLTGKIIGDKGYISQKLFDQLYEKGLQLITKIRKN
                     MKNKLVLMIDKILLKKRGIIA"
     repeat_region   4562. .4849
BASE COUNT     1362 a    778 c   1265 g   1578 t
ORIGIN

  Query Match           100.0%;  Score 486;  DB 1;  Length 4983;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-129;
  Matches  486;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2834 ATGAACAGAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGC 2893

Qy     61 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2894 TGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGT 2953

Qy    121 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2954 GTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCT 3013

Qy    181 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3014 GTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATA 3073

Qy    241 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3074 AAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCA 3133

Qy    301 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3134 GATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAA 3193

Qy    361 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3194 GAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAG 3253

Qy    421 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3254 CAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACA 3313

Qy    481 GAAAAA 486
          ||||||
Db   3314 GAAAAA 3319


RESULT 3
RSU76907
LOCUS       RSU76907                448 bp    DNA     linear   BCT 18-FEB-1998
DEFINITION  Rickettsia sp. 17kDa common-antigen gene, partial cds.
ACCESSION   U76907
VERSION     U76907.1  GI:2894794
KEYWORDS    .
SOURCE      Rickettsia sp.
  ORGANISM  Rickettsia sp.
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Davis,M.J., Ying,Z., Brunner,B.R., Pantoja,A. and Ferwerda,F.H.
  TITLE     Rickettsial relative associated with papaya bunchy top disease
  JOURNAL   Curr. Microbiol. 36 (2), 80-84 (1998)
  MEDLINE   98087556
   PUBMED   9425244
REFERENCE   2  (bases 1 to 448)
  AUTHORS   Ying,Z. and Davis,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1996) Tropical Research and Education Center,
            University of Florida, 18905 SW 280 Street, Homestead, FL 33031,
            USA
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Rickettsia sp."
                     /db_xref="taxon:789"
                     /note="Puerto Rico"
     CDS             <1. .>448
                     /note="putative"
                     /codon_start=3
                     /transl_table=11
                     /product="17 kDa common-antigen"
                     /protein_id="AAC02809.1"
                     /db_xref="GI:2894795"
                     /translation="ALAASMLEACQGPGGMNKQGSGTLIGGTLGGLVGSQFGGGTGRL
                     AAVGAGALLGAILGNQIGAGMDEQDRKLAELTSQRALEAAPSGSSVQWRNPDNGNYGT
                     VTPSKAYKNNTGQYCREYTQTVVVGGKQQKAYGTACRQPDGQWQVV"
BASE COUNT      135 a     81 c    120 g    112 t
ORIGIN

  Query Match           16.3%;  Score 79;  DB 1;  Length 448;
  Best Local Similarity  51.6%;  Pred. No. 8.5e-12;
  Matches  206;  Conservative    0;  Mismatches  190;  Indels    3;  Gaps    1;

Qy     72 CTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCA 131
          | | | |   ||||     ||  |    | ||| |    | | || ||  ||| |||
Db     47 CATGAATAAACAAGGTAGTGGTACGCTAATTGGCGGAACTTTAGGAGGACTTGTTGGTTC 106

Qy    132 GCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGG 191
          |  ||||||   ||||  || ||| ||||   ||      |  |||||  |  | ||||
Db    107 ACAATTTGGTGGTGGTACGGGACGACTTGCTGCGGTAGGAGCAGGTGCGTTACTTGGTGC 166

Qy    192 ATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCA 251
          | |  ||||     |||| |||  |   ||||||  | || |||| || |     | | |
Db    167 AATCCTTGGCAACCAAATTGGTGCAGGTATGGATGAACAAGATAGAAAACTTGCAGAGCT 226

Qy    252 GAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAA 311
          | ||   |||  |    || ||     |   || |  |  | | |  | |     | ||
Db    227 TACTTCTCAAAGAGCTTTAGAAG---CTGCACCAAGCGGTAGCAGTGTACAATGGCGTAA 283

Qy    312 TAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCA 371
          |  | ||| || | |    |    |       |   | |   | |  || | |  |
Db    284 TCCTGATAATGGTAATTATGGTACTGTGACACCAAGCAAAGCTTATAAAAATAATACCGG 343

Qy    372 GCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTA 431
          |||||||||| ||||     ||||  |   |  | |  ||  |  | ||| |   |||
Db    344 TCAATATTGTCGTGAATACACCCAAACAGTTGTAGTAGGTGGAAAACAACAAAAAGCTTA 403

Qy    432 CGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGT 470
          |||||||| ||||| ||||| ||||| |  ||||||||
Db    404 TGGCACTGCCTGCCGTCAACCTGATGGACAATGGCAAGT 442


RESULT 4
RPXX04
LOCUS       RPXX04                237523 bp    DNA     linear   BCT 11-NOV-1998
DEFINITION  Rickettsia prowazekii strain Madrid E, complete genome; segment
            4/4.
ACCESSION   AJ235273 AJ235269
VERSION     AJ235273.1  GI:3861237
KEYWORDS    complete genome.
SOURCE      Rickettsia prowazekii.
  ORGANISM  Rickettsia prowazekii
```

```
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; typhus group.
REFERENCE   1  (bases 1 to 237523)
  AUTHORS   Andersson,S.G., Zomorodipour,A., Andersson,J.O.,
            Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K.,
            Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
  TITLE     The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
  JOURNAL   Nature 396 (6707), 133-140 (1998)
  MEDLINE   99039499
   PUBMED   9823893
REFERENCE   2  (bases 1 to 237523)
  AUTHORS   Andersson,S.G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1998) S.G.E. Andersson,
            Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University
            of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .237523
                     /organism="Rickettsia prowazekii"
                     /strain="Madrid E"
                     /db_xref="taxon:782"
     gene            164. .1123
                     /gene="RP701"
     CDS             164. .1123
                     /gene="RP701"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15137.1"
                     /db_xref="GI:3861238"
                     /translation="MAEPIIFTLLFGNKDRILDRINNQKALNKLHIIPLQVNKESHIS
                     DKERIFSEILEKSYKECKIPIFNIQISNNNVKPIFRAQDLINAQNLNIKTTITFDKYN
                     ALPQNSELKSYWKQIIEKVDHVFFTNEADQNLSIADGIVPKDKTTTITDIRLVTSVFN
                     NLIYNHETNHGLSSTIYKKVDKLIKTAKNQDSKEIIETWPISVDEATNLITAKFGITS
                     EDQIYGLKLEIHEILKDPNNAAENFKKYVSQISRQFKKDLDKDQKNHIDSNFNTQKVI
                     NDSNKDIQVEHITSYVQDKEKQSKPQGFLTRIFNYFKNIVFKQ"
     gene            2074. .3348
                     /gene="RP702"
     CDS             2074. .3348
                     /gene="RP702"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15138.1"
                     /db_xref="GI:3861239"
                     /translation="MTTILVIVIIIMISLSALFAATETAITASSPGKIHKLKIAGNKR
                     AKTVLEVLKKKEKVIGTLLIGNSLINTICTTIATTLFISLLGNNGTIVASGVMAFIII
                     VFAEVVPKAIAVAKPEQLTLKMSSTIVIFLKLFKPINIALDYITKIFCFIFRINLNPQ
                     ISGTEEVRGVIEHYHQEGGVYKSERNMLGGILDIRNMTVSEIMTHRSNIIALNIDLPY
                     ELIIKTLLSGAHTRIPLWQDNRDNIIGILNLKDLLKALYENNNDDKKVDINTLLTPPW
                     FIPDNALVVDQLHAFRERNNHFACVVDEYGTLLGIITLEDVIEEIVGPITDEHDRLNN
                     EIIQESNTEFIIKGTTTIRDINRELDWNLSDEDANTIAGLIIHKIARIPNQGEVIKIF
                     NFKIIILKKIANKIDSVKITVLPMTEETISSE"
     gene            3341. .5353
                     /gene="RP703"
     CDS             3341. .5353
                     /gene="RP703"
                     /codon_start=1
                     /transl_table=11
                     /product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
                     (ccmF)"
                     /protein_id="CAA15139.1"
                     /db_xref="GI:3861240"
                     /translation="MSKIGNFLLLTTMCLSGILTLLIIPWLDYLIKKKHRNNICLKFF
                     PWIPCTSQVMTQAITFYASSLCSIFAFCTLVYAFITSDFSVQNVFLHSSTLKPLIYKI
                     SGSWSSHEGSMLLWFCLLQIVSCCYIFCLEDKMLKFSSIIILSAIQLLFSSFIYFTSN
                     PFHMFSFVPKEGLGLNPMLQDIALSIHPPLLYLGYVSYAVPFTIVCAIFLIAVISRLD
                     NNLPGYFTGYKITQLSDRITSIRLLKIFAGCGILFTTIGISLGSWWAYRELNWGGLWF
                     FDPVENISLFPWLSGIMLHHSIIVTTKTNRMQHWTIILSIITFLLVIFSTFLVRSGFI
                     NSIHAFAFSSERWIYLLVILCIIGIGGLYVIIRLNNCSIEKTIYNTFNWNTHSNCNMT
                     IKEKAILIGNIFFLFSLIILICATLYPTIYALFNDKPIIINEEFFIKNFIIFVIPILC
                     TIGLFTTNSSIKKHIIMLMLSLIITYFILLNIKFSVISILTIMTSIFLMLHNVHYLLV
                     KTNYFRNSLKASSTSMILGHFGVALIAFSITMNSLLQSEMDFIGEVGTSKTFNAFKVT
                     LQNIKFAQGKNYYRQIAEFWLEDHKNNITILKPENRFYIIEQKLSQESDIHSYLLYDL
                     YAVLSNIDGKIIHAKIYYKPMMSFIWIGVILTTAGFVIALIRKNSY"
     gene            complement(7264. .12195)
                     /gene="RP704"
     CDS             complement(7264. .12195)
                     /gene="RP704"
                     /codon_start=1
                     /transl_table=11
                     /product="CELL SURFACE ANTIGEN (sca5)"
                     /protein_id="CAA15140.1"
                     /db_xref="GI:3861241"
                     /translation="MAQKPNFLKKIISAGLVTASTATIVAGFSGVAMGAAMQYNRTTN
                     AAATTFDGIGFDQAAGANIPVAPNSVITANANNPITFNTPNGHLNSLFLDTANDLAVT
                     INEDTTLGFITNIAQQAKFFNFTVAAGKILNITGQGITVQEASNTINAQNALTKVHGG
                     AAINANDLSGLGSITFAAAPSVLEFNLINPTTQEAPLTLGANSKIVNGGNGTLNITNG
                     FIQVSDNTFAGIKTINIDDCQGLMFNSTPDAANTLNLQVGGNTINFNGIDGTGKLVLV
                     SKNGAATEFNVTGTLGGNLKGIIELNTAAVAGKLISQGGAANAVIGTDNGAGRAAGFI
                     VSVDNGNAATISGQVYAKNMVIQSANAGGQVTFEHIVDVGLGGTTNFKTADSKVIITE
                     NSNFGSTNFGNLDTQIVVPDTKILKGNFIGDVKNNGNTAGVITFNANGALVSASTDPN
                     IAVTNINAIEAEGAGVVELSGIHIAELRLGNGGSIFKLADGTVINGPVNQNALMNNNA
                     LAAGSIQLDGSAIITGDIGNGGVNAALQHITLANDASKILALDGANIIGANVGGAIHF
                     QANGGTIKLTNTQNNIVVNFDLDITTDKTGVVDASSLTNNQTLTINGSIGTVVANTKT
                     LAQLNIGSSKTILNAGDVAINELVIENNGSVQLNHNTYLITKTINAANQGQIIVAADP
                     LNTNTTLADGTNLGSAENPLSTIHFATKAANADSILNVGKGVNLYANNITTNDANVGS
                     LHFRSGGTSIVSGTVGGQQGHKLNNLILDNGTTVKFLGDTTFNGGTKIEGKSILQISN
                     NYTTDHVESADNTGTLEFVNTDPITVTLNKQGAYFGVLKQVIISGPGNIVFNEIGNVG
                     IVHGIAANSISFENASLGTSLFLPSGTPLDVLTIKSTVGNGTVDNFNAPIVVVSGIDS
                     MINNGQIIGDKKNIIALSLGSDNSITVNANTLYSGIRTTKNNQGTVTLSGGMPNNPGT
                     IYGLGLENGSPKLKQVTFTTDYNNLGSIIANNVTINDYVTLTTGGIAGTDFDAKITLG
                     SVNGNANVRFVDSTFSDPRSMIVATQANKGTVTYLGNALVSNIGSLDTPVASVRFTGN
                     DSGAGLQGNIYSQNIDFGTYNLTILNSNVILGGGTTAINGEIDLLTNNLIFANGTSTW
                     GDNTSISTTLNVSSGNIGQVVIAEDAQVNATTTGTTTIKIQDNANANFSGTQAYTLIQ
                     GGARFNGTLGAPNFAVTGSNIFVKYELIRDSNQDYVLTRTNDVLNVVTTAVGNSAIAN
                     APGVSQNISRCLESTNTAAYNNMLLAKDPSDVATFVGAIATDTSAAVTTVNLNDTQKT
                     QDLLSNRLGTLRYLSNAETSDVAGSATGAVSSGDEAEVSYGVWAKPFYNIAEQDKKGG
                     IAGYKAKTTGVVVGLDTLASDNLMIGAAIGITKTDIKHQDYKKGDKTDINGLSFSLYG
                     SQQLVKNFFAQGNAIFTLNKVKSKSQRYFFESNGKMSKQIAAGNYDNMTFGGNLIFGY
                     DYNAMPNVLVTPMAGLSYLKSSNENYKETGTTVANKRINSKFSDRVDLIVGAKVAGST
                     VNITDIVIYPEIHSFVVHKVNGKLSNSQSMLDGQTAPFISQPDRTAKTSYNIGLSANI
                     KSDAKMEYGIGYDFNSASKYTAHQGTLKVRVNF"
     gene            complement(12953. .13216)
                     /gene="RP705"
     misc_feature    complement(12953. .13216)
                     /gene="RP705"
                     /note="Pseudogene that lacks ATG or GTG initiation codon.
                     Homology to guanosine-3,5-bis(diphosphate)
                     3-pyrophosphohydrolase."
     gene            complement(13563. .14507)
                     /gene="RP706"
     CDS             complement(13563. .14507)
                     /gene="RP706"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15141.1"
                     /db_xref="GI:3861242"
                     /translation="MKIRQTAKPVIIGISGPELTDAERELFEEHNPLGIILFRRNIRK
                     NEKGEQDKAALIKLIADIKEVLGDNTIIAIDQEGGMVKRLIAPTFYDAPAAQTFTELQ
                     TCKYNYSTIAKELREVGINLDFAPVADLIHDGADKIISDRSFGKEPEIVVPLFLSAIA
                     GLQEEKVTACIKHIPGHGRATVDSHIELPIIDTSLKTLEDTDFKVFKELAKYDYIKLA
                     MTAHVIYTALDPDNPATLSQIVIDYIKSNIGFKGLIISDTIEMQALSGSMADITKGAL
                     DAGVDIVLECTGEFNKMSEVLGSVAENSIDKFSDLFIS"
     gene            complement(14552. .15025)
                     /gene="RP707"
     CDS             complement(14552. .15025)
                     /gene="RP707"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA15142.1"
                     /db_xref="GI:3861243"
                     /translation="MINNKLTKKYYVASEVIRILNIALHKLRYLETKIPGLSHYKINN
```

```
                RKYYTAHDIDLLQKFLFVSRREEYLEVRGYHRSDKDYLANFSNIHCASDTSRQNEFQG
                KNSECILNLHVQQRQDSQSAVVLSLMDNTVSSRIDILLTNFHNLSIQIKKILLIS"
     gene       complement(15220. .15507)
                /gene="RP708"
     CDS        complement(15220. .15507)
                /gene="RP708"
                /codon_start=1
                /transl_table=11
                /product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (himA)"
                /protein_id="CAA15143.1"
                /db_xref="GI:3861244"
                /translation="MTITKNKISLMLSSKLGFSNNLCEAIVNTVFSNILEIAKVQKLT
                LKNFGSFEVKQKTPRPGINFHTKSPVMIASKKNLRFTPSEKLKALINKSML"
     gene       18405. .18701
                /gene="RP709"
     CDS        18405. .18701
                /gene="RP709"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="CAA15144.1"
                /db_xref="GI:3861245"
                /translation="MAVGRQYIEDHTEQLHKLCDNLVKYIELETKAITHEKAAEDTEV
                VIEVSLDLDNSSDDASIINKKAITVTSDDIANSNEDAARNTNKSHGDVLRCYYQ"
     gene       complement(19685. .20789)
                /gene="RP710"
     misc_feature complement(19685. .20789)
                /gene="RP710"
                /note="Pseudogene with two authentic frameshifts and four
                authentic in-frame termination codons. Homology to
                transposase."
     gene       complement(21112. .21608)
                /gene="RP711"
     misc_feature complement(21112. .21608)
                /gene="RP711"
                /note="Pseudogene with one authentic frameshift. Homology
                to resolvase."
     gene       complement(22602. .23066)
                /gene="RP712"
     CDS        complement(22602. .23066)
                /gene="RP712"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="CAA15145.1"
                /db_xref="GI:3861246"
```

Query Match 14.5%; Score 70.6; DB 1; Length 237523;
Best Local Similarity 50.6%; Pred. No. 6.4e-09;
Matches 197; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

```
Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
           ||||   |||  |      ||| |  |  |  || |   |   |||    |  || |||
Db 172624 CAAGGTACCGGTACACTTCTTGGTGGCGCAGGAGGTGCATTACTTGGTTCTCAATTCGGT 172683

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           ||||||  || | | |||     |    ||  || ||  |  | || | | |  |||||
Db 172684 CAAGGTAAAGGACAACTTGTCGGAGTAGGTGTAGGCGCATTACTTGGGGCAGTTCTTGGT 172743

Qy     202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAA 261
           |||||||   |   |||||| |||||||| || |   ||| | |  |  |   ||
Db 172744 GGACAAATCGGCGCAAGTATGGATGAGCAGGATAGAAGACTTCTAGAACTAACATCACAA 172803

Qy     262 AAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGT 321
           | |    || |     |   |   ||    ||||   |||     |||||    |||
Db 172804 AGAGCTTTAGAATCTGCACCTAGCGGTA---GTAACATAGAATGGCGCAATCCAGATAAC 172860

Qy     322 GTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGT 381
           |   | |  |           |    | | | | |     |  ||      ||||||||
Db 172861 GGCAATCATGGTTACGTCACACCTAATAAAACTTATAGGAACAGTGCAGGTCAATATTGC 172920

Qy     382 CGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCA 441
           || ||||     ||||  |  || || |  ||  |  | ||| | |  || || | ||||
```

```
Db 172921 CGTGAATACACTCAAACAGTTATAATAGGCGGAAAACAACAAAAAACATATGGTAATGCA 172980

Qy     442 TGCCGGCAACCGGATGGTCGTTGGCAAGT 470
           ||||| ||||| ||||| |  ||||||||
Db 172981 TGCCGTCAACCTGATGGGCAATGGCAAGT 173009


RESULT 5
AF031534
LOCUS       AF031534               416 bp    DNA     linear   BCT 05-MAR-2002
DEFINITION  Rickettsia cooleyi 17 kDa antigen gene, partial cds.
ACCESSION   AF031534
VERSION     AF031534.1  GI:2654012
KEYWORDS    .
SOURCE      Rickettsia cooleyi.
  ORGANISM  Rickettsia cooleyi
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Billings,A.N., Teltow,G.J., Weaver,S.C. and Walker,D.H.
  TITLE     Molecular characterization of a novel Rickettsia species from
            Ixodes scapularis in Texas
  JOURNAL   Emerging Infect. Dis. 4 (2), 305-309 (1998)
  MEDLINE   98284224
   PUBMED   9621204
REFERENCE   2  (bases 1 to 416)
  AUTHORS   Billings,A.N., Teltow,G.J. and Walker,D.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1997) Pathology, UTMB at Galveston, 301
            University Blvd., Galveston, TX 77555-0609, USA
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Rickettsia cooleyi"
                     /db_xref="taxon:69410"
     CDS             <1. .415
                     /codon_start=2
                     /transl_table=11
                     /product="17 kDa antigen"
                     /protein_id="AAB95267.1"
                     /db_xref="GI:2654013"
                     /translation="GSYIGPRGMNKQGTGTLLGGAGGALLGSQFGKGKGQLVGVGVGA
                     LLGAVLGGQIGAGMDEQDRRLAELTSQRALEAAPSGSSTEWRNPDNGNYGYVTPNKTY
                     RNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDEQ"
BASE COUNT      133 a     80 c    108 g     95 t
ORIGIN
```

Query Match 14.1%; Score 68.4; DB 1; Length 416;
Best Local Similarity 49.5%; Pred. No. 9.9e-09;
Matches 205; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

```
Qy      51 TTTAGTTGGCTGTGCCCAGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGT 110
           || |  |  | || | |       | | |   ||||     ||| |       ||| | ||
Db       4 TTCATATATCGGTCCACGCGGTATGAATAAACAAGGCACAGGAACACTTCTTGGCGGTGC 63

Qy     111 TGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCAT 170
           |  || |   |  ||||   |  || ||||||||||  || |   |||    |
Db      64 CGGCGGTGCATTACTTGGCTCTCAATTCGGTAAAGGTAAAGGACAGCTTGTCGGAGTAGG 123

Qy     171 TGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCA 230
           ||  |||||  |  | || | | |  |||||    ||||||||  |   |||||| ||||
Db     124 TGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTATGGATGAGCA 183

Qy     231 GGATAAAATAAAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTG 290
           ||||| || |     | | |  | |  |||  |    || ||    ||  |  | |  |
Db     184 GGATAGAAGACTTGCAGAACTCACCTCACAAAGAGCTTTAGAAG---CAGCTCCTAGCGG 240

Qy     291 GCGTAATCCAGATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTA 350
           |||  | ||||     | |||  | |||  |    |    |    |      |    | |
Db     241 TAGTAGTACAGAATGGCGTAATCCTGATAACGGCAATTACGGTTATGTAACACCTAATAA 300

Qy     351 CAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGC 410
           | | |   | |  |       |||||||| || || | |   ||||  |   | || |
Db     301 AACTTATAGAAATAGCACTGGTCAATATTGCCGTGAGTATACTCAAACAGTTGTAATAGG 360
```

```
Qy    411 AGGGCAGAAGCAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTG 464
          ||  |  | ||| |    ||||| | ||||||||| ||||| || |  |  ||
Db    361 CGGAAAACAACAAAAAGCATACGGTAATGCATGCCGCCAACCTGACGAACAATG 414


RESULT 6
RIRANT17KB
LOCUS       RIRANT17KB              535 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.typhi 17K genus-common antigen gene, complete cds.
ACCESSION   M28481
VERSION     M28481.1  GI:152459
KEYWORDS    antigen.
SOURCE      R.typhi DNA.
  ORGANISM  Rickettsia typhi
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; typhus group.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Anderson,B.E. and Tzianabos,T.
  TITLE     Comparative sequence analysis of a genus-common rickettsial antigen
            gene
  JOURNAL   J. Bacteriol. 171 (9), 5199-5201 (1989)
  MEDLINE   89359171
   PUBMED   2768201
COMMENT     Draft entry and printed copy of sequence for [1] kindly submitted
            by B.Anderson, 27-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Rickettsia typhi"
                     /db_xref="taxon:785"
     -35_signal      1. .6
     -10_signal      21. .27
     mRNA            37. .>535
                     /product="17 kD mRNA"
     CDS             56. .535
                     /note="17 kD antigen precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26377.1"
                     /db_xref="GI:152460"
                     /translation="MKLLSKVMILALAASMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGHGKGQLVGVGVGALLGAVLGGQIGASLDEQDRKLLELTSQRALESAPSGSNIEW
                     RNPDNGNHGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQTTYGNACRQPDGQWQVVN"
     sig_peptide     56. .115
                     /note="17 kD antigen signal peptide"
     mat_peptide     116. .532
                     /product="17 kD antigen"
BASE COUNT      182 a     93 c    122 g    138 t
ORIGIN

  Query Match             13.5%;  Score 65.8;  DB 1;  Length 535;
  Best Local Similarity   49.9%;  Pred. No. 5.8e-08;
  Matches  194;  Conservative    0;  Mismatches  192;  Indels    3;  Gaps    1;

Qy     82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
          ||||     ||| |       ||| |  |  |  || |    |   |||     |  || |||
Db    140 CAAGGGACTGGAACACTTCTTGGTGGCGCCGGCGGTGCATTACTTGGTTCTCAATTCGGT 199

Qy    142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
          | ||||  || | | |||     |    ||  || ||  |  | || | | |  |||||
Db    200 CACGGTAAAGGACAACTTGTCGGAGTAGGTGTAGGCGCATTACTTGGGGCAGTTCTTGGT 259

Qy    202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAA 261
          ||||||||  |    ||||| |||||||||| || |   ||| | |  |  |   ||
Db    260 GGACAAATCGGTGCAAGTCTGGATGAGCAGGATAGAAAACTTCTAGAACTAACATCACAA 319

Qy    262 AAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGT 321
          |  |    || |      |  | |  |  |||| |||     |||||    ||| |
Db    320 AGAGCTTTAGAATCT---GCTCCTAGCGGTAGTAACATAGAATGGCGCAATCCAGATAAT 376

Qy    322 GTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGT 381
          |  | |  |           |    | | | | |     |  ||      ||||||||
Db    377 GGCAATCATGGTTACGTAACACCTAATAAAACTTATAGGAACAGTACAGGTCAATATTGC 436

Qy    382 CGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCA 441
          || ||||     ||||  |   | || |  ||  |  | |||   |  || || | ||||
Db    437 CGTGAATACACTCAAACAGTTGTAATAGGGGGAAAACAACAAACAACATATGGAAATGCA 496

Qy    442 TGCCGGCAACCGGATGGTCGTTGGCAAGT 470
          ||||| ||||| || || |  |||||||
Db    497 TGCCGCCAACCTGACGGACAATGGCAAGT 525


RESULT 7
RIR17KGCA
LOCUS       RIR17KGCA               537 bp    DNA     linear   BCT 04-FEB-1999
DEFINITION  Rickettsia japonica 17K genus-common antigen gene, complete cds.
ACCESSION   D16515
VERSION     D16515.1  GI:695412
KEYWORDS    17K genus-common antigen; 17k dalton protein.
SOURCE      Rickettsia japonica (sub_species:YH) DNA.
  ORGANISM  Rickettsia japonica
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Furuya,Y., Katayama,T., Yoshida,Y. and Kaiho,I.
  TITLE     Specific amplification of Rickettsia japonica DNA from clinical
            specimens by PCR
  JOURNAL   J. Clin. Microbiol. 33 (2), 487-489 (1995)
  MEDLINE   95229950
REFERENCE   2  (bases 1 to 537)
  AUTHORS   Yoshida,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1993) Yoshiya Yoshida, Kanagawa Prefectural
            Pabulic Health Laboratoty; 52-2 Nakao-cho, Asahi-ku, Yokohama,
            Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Rickettsia japonica"
                     /sub_species="YH"
                     /db_xref="taxon:35790"
                     /note="propagated in Vero cells"
     CDS             58. .537
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA03965.1"
                     /db_xref="GI:1435056"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGTGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGQWQVVN"
BASE COUNT      176 a     92 c    129 g    140 t
ORIGIN

  Query Match             13.5%;  Score 65.8;  DB 1;  Length 537;
  Best Local Similarity   49.9%;  Pred. No. 5.8e-08;
  Matches  194;  Conservative    0;  Mismatches  192;  Indels    3;  Gaps    1;

Qy     82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
          ||||     ||| |       ||| | || || |||||   |   |||    | || |||
Db    142 CAAGGTACAGGAACACTTCTTGGCGGTGCTGGTGGCGCATTACTTGGTTCTCAATTCGGT 201

Qy    142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
          || || |  || |   |||      |    ||  ||||  |  | || | | |  |||||
Db    202 AAGGGCACAGGACAGCTTGTGGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGT 261

Qy    202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAA 261
          ||||||||  |   |||||| |||||||||| || |     | | |  |  |   |
Db    262 GGACAAATCGGTGCAGGTATGGATGAGCAGGATAGAAGACTTGCAGAGCTTACCTCACAG 321

Qy    262 AAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGT 321
          |  |    || |      ||   |  || |  ||||  |||     | |||    |||
Db    322 AGAGCTTTAGAAAC---AGCTCCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAAC 378

Qy    322 GTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGT 381
          |  |   |           |    | | | | |  | |  |       ||||||||
Db    379 GGCAATCATGGTTACGTAACACCTAATAAAACTTATAGGAACAGTACAGGTCAATATTGC 438
```

```
Db    379 GGCAATTACGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGC 438

Qy    382 CGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCA 441
          || || |     ||||  |   | || |  ||  |  | ||| |    ||||| | |||
Db    439 CGTGAGTACACTCAAACAGTTGTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCG 498

Qy    442 TGCCGGCAACCGGATGGTCGTTGGCAAGT 470
          ||||| ||||| || || |  ||||||||
Db    499 TGCCGCCAACCTGACGGACAATGGCAAGT 527


RESULT 8
RIRANT17KA
LOCUS       RIRANT17KA               539 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.conorii 17K genus-common antigen gene, complete cds.
ACCESSION   M28480
VERSION     M28480.1  GI:152457
KEYWORDS    antigen.
SOURCE      R.conorii DNA.
  ORGANISM  Rickettsia conorii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Anderson,B.E. and Tzianabos,T.
  TITLE     Comparative sequence analysis of a genus-common rickettsial antigen
            gene
  JOURNAL   J. Bacteriol. 171 (9), 5199-5201 (1989)
  MEDLINE   89359171
   PUBMED   2768201
COMMENT     Draft entry and printed copy of sequence for [1] kindly submitted
            by B.Anderson, 27-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Rickettsia conorii"
                     /db_xref="taxon:781"
     -35_signal      1. .6
     -10_signal      21. .27
     mRNA            38. .>539
                     /product="17 kD mRNA"
     CDS             60. .539
                     /note="17 kD antigen precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26376.1"
                     /db_xref="GI:152458"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGKGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGQWQVVN"
     sig_peptide     60. .119
                     /note="17 kD antigen signal peptide"
     mat_peptide     120. .536
                     /product="17 kD antigen"
BASE COUNT      179 a     93 c    125 g    142 t
ORIGIN

  Query Match           13.3%;  Score 64.6;  DB 1;  Length 539;
  Best Local Similarity   49.4%;  Pred. No. 1.3e-07;
  Matches  197;  Conservative    0;  Mismatches  199;  Indels    3;  Gaps    1;

Qy     72 CTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCA 131
          | | | |   ||||     ||| |      ||| | || ||  ||||   |   |||
Db    134 CATGAATAAACAAGGTACAGGAACACTTCTTGGCGGTGCTGGCGGCGCATTACTTGGTTC 193

Qy    132 GCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGG 191
            |  || ||||| || |  || |   |||     |    ||  |||||  |  | || |
Db    194 TCAATTCGGTAAGGGCAAAGGACAGCTTGTTGGAGTAGGTGTAGGTGCATTACTTGGAGC 253

Qy    192 ATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCA 251
          | |  |||||    ||||||||  |   |||||| | ||||||| || |     | | |
Db    254 AGTTCTTGGTGGACAAATCGGTGCAGGTATGGATGAACAGGATAGAAGACTTGCAGAGCT 313

Qy    252 GAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAA 311
          |  |   | |  |    || |     ||   |  || |  ||||   |||     | ||

Db    314 TACCTCACAGAGAGCTTTAGAAAC---AGCTCCTAGTGGTAGTAACGTAGAATGGCGTAA 370

Qy    312 TAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCA 371
          |    |||  |   |    |            |    | | | | |   | |  |
Db    371 TCCGGATAACGGCAATTACGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGG 430

Qy    372 GCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTA 431
          ||||||||  || || |     ||||  |   | || |  ||  |  | ||| |    ||
Db    431 TCAATATTGCCGTGAGTACACTCAAACAGTTGTAATAGGCGGAAAACAACAAAAAGCATA 490

Qy    432 CGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGT 470
          ||| | |||||||||| ||||| || || |  ||||||||
Db    491 CGGTAATGCATGCCGCCAACCTGACGGACAATGGCAAGT 529


RESULT 9
AE008675
LOCUS       AE008675              10127 bp    DNA     linear   BCT 14-SEP-2001
DEFINITION  Rickettsia conorii Malish 7, section 107 of 114 of the complete
            genome.
ACCESSION   AE008675 AE006914
VERSION     AE008675.1  GI:15620413
KEYWORDS    .
SOURCE      Rickettsia conorii.
  ORGANISM  Rickettsia conorii
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (sites)
  AUTHORS   Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E.,
            Raoult,D. and Claverie,J.M.
  TITLE     Selfish DNA in protein-coding genes of Rickettsia
  JOURNAL   Science 290 (5490), 347-350 (2000)
  MEDLINE   20485642
   PUBMED   11030655
REFERENCE   2  (sites)
  AUTHORS   Ogata,H., Audic,S. and Claverie,J.-M.
  TITLE     Selfish DNA and the origin of genes
  JOURNAL   Science 291 (5502), 252-253 (2001)
REFERENCE   3  (bases 1 to 10127)
  AUTHORS   Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
            Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
            Raoult,D.
  TITLE     Mechanisms of evolution in Rickettsia conorii and R. prowazekii
  JOURNAL   Science 293 (5537), 2093-2098 (2001)
  MEDLINE   21442074
   PUBMED   11557893
REFERENCE   4  (bases 1 to 10127)
  AUTHORS   Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V.,
            Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. and
            Raoult,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
            CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean
            Moulin, Marseille Cedex 05 13385, France
COMMENT     A public version of R. conorii genome database is accessible at
            http://igs-server.cnrs-mrs.fr/. The database intends to provide
            updated data. Annotation of the genome is an ongoing task whose
            goal is to make the genome sequence more useful. Comments to the
            authors are appreciated.
FEATURES             Location/Qualifiers
     source          1. .10127
                     /organism="Rickettsia conorii"
                     /strain="Malish 7"
                     /db_xref="taxon:781"
     gene            111. .890
                     /gene="map"
                     /note="RC1276"
     CDS             111. .890
                     /gene="map"
                     /codon_start=1
                     /transl_table=11
                     /product="methionine aminopeptidase [EC:3.4.11.18]"
                     /protein_id="AAL03814.1"
```

```
                     /db_xref="GI:15620414"
                     /translation="MTIKIHTGKDFIKMRAAGKLAAETLDFITDHVKPNVTTNSLNGL
                     CHNFITSHNAVPAPLNYKGFPKSICTSINHVVCHGIPNDKPLKNGDIVNIDVTVILDG
                     WYGDTSRMYYVGDVAIKPKRLIQVTYDAMMKGIEVVRPGAKLGDIGHAIQSYAEKHNY
                     SVVRDYTGHGIGRVFHDEPSILNYGRSGTGLTLEEGMFFTVEPMINAGNYDTILSKLD
                     GWTVTTRDKSLSAQFEHTIGVTKDGFEIFTLSPKKLDYPPY"
repeat_region        complement(977. .1011)
                     /note="REP03, repeated element"
repeat_region        complement(1064. .1083)
                     /note="REP02, RS3-like repeat"
gene                 1141. .1452
                     /gene="RC1277"
CDS                  1141. .1452
                     /gene="RC1277"
                     /codon_start=1
                     /transl_table=11
                     /product="similarity to DNA repair protein (RadC)"
                     /protein_id="AAL03815.1"
                     /db_xref="GI:15620415"
                     /translation="MLENNLLEFSNDEIIDTPHYIGHRKRVKERFVAAGAEYFSDYEL
                     LEVMLFSAIPRKDIKPLAKKLLDHFDITDLINLDKERLLSIKGTNENLYINFALRELI
                     N"
gene                 1457. .1825
                     /gene="RC1278"
CDS                  1457. .1825
                     /gene="RC1278"
                     /codon_start=1
                     /transl_table=11
                     /product="similarity to DNA repair protein (RadC)"
                     /protein_id="AAL03816.1"
                     /db_xref="GI:15620416"
                     /translation="MKQKIINKNINASWSSWIDYLKVNMGNMRLEQFRILFLNKKNIL
                     IADEVLSQGTIDQAAVYPREIIKRALFNEASSLILVHHHPSGSPEPSKANIHMTNKIV
                     EKCQTINIIVLNLICSYEYK"
gene                 1812. .2795
                     /gene="mraY2"
                     /note="RC1279"
CDS                  1812. .2795
                     /gene="mraY2"
                     /codon_start=1
                     /transl_table=11
                     /product="undecaprenyl-phosphate
                     alpha-N-acetylglucosaminyltransferase [EC:2.4.1.-]"
                     /protein_id="AAL03817.1"
                     /db_xref="GI:15620417"
                     /translation="MNINSYSTLFIFSFISTAILTYILTKYLPTIGLVDIPSNRRSHD
                     KITPRGGGLAIVFVVMIALSGFEYITSNNFVNSIKILPLLLAIASISFLDDLKAVPIL
                     IRLIVHLICAAFAIFLFSQLNSVHILIYSILVIALSGFINIYNFMDGIDGMSCVESIH
                     LSSTMLILCFLQFPTINNPYFITSVNVIILGGSCGFLIFNWHPAKIFLGDVGSISLGF
                     LTGLCLLLLALTSTNLFIACAIASLYYITDAVLTILIRLLNKEKIWQPHLKHFFQKAV
                     QKGKSHKQVVSIIAVCNIFLMIISVTSLHFPVISIMLALAVITLTVQSLLK"
gene                 2792. .3115
                     /gene="RC1280"
CDS                  2792. .3115
                     /gene="RC1280"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL03818.1"
                     /db_xref="GI:15620418"
                     /translation="MKLIKEDIDKIVRHIFAKQHPLLPEIMINWNKIVGFNFSTKALP
                     LKITTYTYKKQKINTLFIQAEDNATAAELPYYQDIILERIKIYLGFAAIHQMNVTFYK
                     GKKSL"
repeat_region        complement(3198. .3227)
                     /note="REP02, RS3-like repeat"
repeat_region        3313. .3353
                     /note="REP01, repeated element"
gene                 complement(3509. .4258)
                     /gene="RC1281"
CDS                  complement(3509. .4258)
                     /gene="RC1281"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL03819.1"
                     /db_xref="GI:15620419"
                     /translation="MKKLLLIAAASTALLTSGLSFADCDMNSSVDSSTNSSMSSSVEN
                     QWYLKLNAGGVIFNKTKPKGADFKLNNIKSNIKSNTGFTGEIGAGYYIMDNLRTDLTI
                     GTVASSHLKKSKTYPDGNSFSVKNKPTIVSVLLNGYVDFVDLSMFKVFAGAGVGAAFV
                     KEKIHSKDIKGGVTDTFNGTTKNKTNFAYQLSLGTSFEVAQGVKAELVYSWRDYGKTK
                     NTTKTINGDKVKFGGTHYKGHNLMAGLRFDM"
repeat_region        4438. .4463
                     /note="REP03, repeated element"
gene                 complement(4505. .5188)
                     /gene="RC1282"
CDS                  complement(4505. .5188)
                     /gene="RC1282"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL03820.1"
                     /db_xref="GI:15620420"
                     /translation="MLIMKKLLLIAATSATMLSSSVSFAEGMDHEWYLRIDTGAAMFN
                     EEKDKATGVKLKSNTTVPVALGIGYYISENFRADLTLGTIIGGKLKKSGAATNAPFTR
                     TNISASHKPTITRLLINGYVDLTNFDMFDVFAGAGVGSALVKEKITYNGITGLSSNTK
                     NRTNISYKLTLGTSAQIADGVKVELAYSWIDDGRTKSKNVIYQGTSVPTGGMHYQSHN
                     LTAGIRFDI"
repeat_region        complement(5338. .5508)
                     /note="REP06, A and T rich repeated element"
gene                 complement(5478. .5828)
                     /gene="fdxA"
                     /note="RC1283"
CDS                  complement(5478. .5828)
                     /gene="fdxA"
                     /codon_start=1
                     /transl_table=11
                     /product="ferredoxin"
                     /protein_id="AAL03821.1"
                     /db_xref="GI:15620421"
                     /translation="MTYVVTDECVKCKYTDCVEVCPVDCFYEGEREDDFMLVINPDEC
                     IDCGVCVPDCPIGAIKPESPGLIEWVERAKDFIENKGWKNITKKKTALPDADKFKDEK
                     DKFNKYIGSVKIAN"
gene                 complement(5833. .6525)
                     /gene="ccmC"
                     /note="RC1284"
CDS                  complement(5833. .6525)
                     /gene="ccmC"
                     /codon_start=1
                     /transl_table=11
                     /product="heme exporter protein C"
                     /protein_id="AAL03822.1"
                     /db_xref="GI:15620422"
                     /translation="MLKLLTPYYFNKLSKVIIPILAISTCVMMIIGLYLALIVSPVDY
                     QQGDFVRIMYVHVPASWMALGVYVFMAVCSFSYLVWKTTISYLLAVASSYVGATFTLI
                     SLVTGSLWGKPIWGTWWVWDARLTSMLILFLLYLSYIIIVNSADNIRKAQNPSSIIAL
                     IGLINIPIVKFSVNIWYSLHQPASVLRLGSPTIHSSMLKPLIIMFISFILYFLLVLLL
                     RTAILIDKIKNR"
gene                 complement(6540. .7046)
                     /gene="RC1285"
CDS                  complement(6540. .7046)
                     /gene="RC1285"
                     /codon_start=1
                     /transl_table=11
                     /product="cytosine deaminase [EC:3.5.4.1]"
                     /protein_id="AAL03823.1"
                     /db_xref="GI:15620423"
                     /translation="MNILLDKEANFNNFFMEQALKQAKIAFDKNEVPVGAVVVDRLHQ
                     KIIASTHNNTEEKNNALYHAEIIAINEACNLISSKNLNDYDIYVTLEPCAMCAAAIAH
                     SRLKRLFYGASDSKHGVVESNLRYFNSSACFHRPEIYSGILAEDSGLLMKEFFKRIRT
                     VISSHRMT"
gene                 complement(7027. .7947)
                     /gene="p34"
                     /note="RC1286"
CDS                  complement(7027. .7947)
```

Query Match 13.3%; Score 64.6; DB 1; Length 10127;

```
  Best Local Similarity   49.4%;  Pred. No. 2.1e-07;
  Matches  197;  Conservative    0;  Mismatches  199;  Indels    3;  Gaps    1;

Qy      72 CTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCA 131
           | | | |   ||||     ||| |       ||| | || ||  ||||    |   |||
Db    8208 CATGAATAAACAAGGTACAGGAACACTTCTTGGCGGTGCTGGCGGCGCATTACTTGGTTC 8267

Qy     132 GCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGG 191
           |  || ||||| || |  || |    |||      |    ||  |||||  |  | || |
Db    8268 TCAATTCGGTAAGGGCAAAGGACAGCTTGTTGGAGTAGGTGTAGGTGCATTACTTGGAGC 8327

Qy     192 ATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCA 251
           | |  |||||    ||||||||  |   |||||| | ||||||| || |     | | |
Db    8328 AGTTCTTGGTGGACAAATCGGTGCAGGTATGGATGAACAGGATAGAAGACTTGCAGAGCT 8387

Qy     252 GAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAA 311
           |  |   | |  |    || |     ||   |  || |  ||||   |||     | ||
Db    8388 TACCTCACAGAGAGCTTTAGAAAC---AGCTCCTAGTGGTAGTAACGTAGAATGGCGTAA 8444

Qy     312 TAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCA 371
           |    |||  |   |   |      |    | | | | |   | |  |
Db    8445 TCCGGATAACGGCAATTACGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGG 8504

Qy     372 GCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTA 431
           |||||||| || || |     ||||  |   | || |  ||  |  | ||| |    ||
Db    8505 TCAATATTGCCGTGAGTACACTCAAACAGTTGTAATAGGCGGAAAACAACAAAAAGCATA 8564

Qy     432 CGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGT 470
           ||| | |||||||| ||||| || || |  ||||||||
Db    8565 CGGTAATGCATGCCGCCAACCTGACGGACAATGGCAAGT 8603


RESULT 10
RIRANT17KD
LOCUS       RIRANT17KD               539 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.rickettsia 17K genus-common antigen gene, complete cds.
ACCESSION   M28479
VERSION     M28479.1  GI:152463
KEYWORDS    antigen.
SOURCE      R.rickettsia DNA.
  ORGANISM  Rickettsia rickettsii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Anderson,B.E. and Tzianabos,T.
  TITLE     Comparative sequence analysis of a genus-common rickettsial antigen
            gene
  JOURNAL   J. Bacteriol. 171 (9), 5199-5201 (1989)
  MEDLINE   89359171
   PUBMED   2768201
COMMENT     Draft entry and printed copy of sequence for [1] kindly submitted
            by B.Anderson, 27-OCT-1989.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Rickettsia rickettsii"
                     /db_xref="taxon:783"
     -35_signal      1. .6
     -10_signal      21. .27
     mRNA            37. .>539
                     /product="17 kD mRNA"
     CDS             60. .539
                     /note="17 kD antigen precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26379.1"
                     /db_xref="GI:152464"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGKGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGQWQVVN"
     sig_peptide     60. .119
                     /note="17 kD antigen signal peptide"
     mat_peptide     120. .536
                     /product="17 kD antigen"

BASE COUNT      179 a     92 c    125 g    143 t
ORIGIN

  Query Match           13.2%;  Score 64.2;  DB 1;  Length 539;
  Best Local Similarity   49.6%;  Pred. No. 1.7e-07;
  Matches  193;  Conservative    0;  Mismatches  193;  Indels    3;  Gaps    1;

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
           ||||     ||| |       ||| | || ||  ||||    |   |||     |  || |||
Db    144 CAAGGTACAGGAACACTTCTTGGCGGTGCTGGCGGCGCATTACTTGGTTCTCAATTCGGT 203

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           || || |  || |   |||      |     ||  |||||  |  | || | | |  |||||
Db    204 AAGGGCAAAGGACAGCTTGTTGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGT 263

Qy     202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAA 261
           ||||||||  |   |||||| | ||||||| || |     | | |  |  |   |
Db    264 GGACAAATCGGTGCAGGTATGGATGAACAGGATAGAAGACTTGCAGAGCTTACCTCACAG 323

Qy     262 AAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGT 321
           |  |    || |     ||   |  || |  ||||   |||     | |||    |||
Db    324 AGAGCTTTAGAAAC---AGCTCCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAAC 380

Qy     322 GTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGT 381
           |  |    |              |    | | | | |  | |      ||||||||
Db    381 GGCAATTACGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGC 440

Qy     382 CGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCA 441
           || || |     ||||  |   | || |  ||  |  | ||| |    ||||| | ||||
Db    441 CGTGAGTACACTCAAACAGTTGTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCA 500

Qy     442 TGCCGGCAACCGGATGGTCGTTGGCAAGT 470
           ||||| ||||| || || |  ||||||||
Db    501 TGCCGCCAACCTGACGGACAATGGCAAGT 529


RESULT 11
RIRANTRR
LOCUS       RIRANTRR                 620 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.rickettsii 17K antigen gene, complete cds.
ACCESSION   M16486
VERSION     M16486.1  GI:152467
KEYWORDS    antigen.
SOURCE      R.rickettsii (strain Sheila Smith) DNA, clone pSC2.
  ORGANISM  Rickettsia rickettsii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Anderson,B.E., Regnery,R.L., Carlone,G.M., Tzianabos,T.,
            McDade,J.E., Fu,Z.Y. and Bellini,W.J.
  TITLE     Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
            rickettsii
  JOURNAL   J. Bacteriol. 169 (6), 2385-2390 (1987)
  MEDLINE   87222152
   PUBMED   3108232
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by B.E.Anderson, 23-OCT-1987.
            -35 and -10 regions are located at positions 12-17 and 33-38.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Rickettsia rickettsii"
                     /db_xref="taxon:783"
     CDS             70. .549
                     /note="17kd antigen precursor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26381.1"
                     /db_xref="GI:152468"
                     /translation="MKLLSKIMIIALATSMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGKGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEW
                     RNPDNGNYGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGDACRQPDEQWQVVN"
     sig_peptide     70. .126
                     /note="17kd antigen signal peptide"
```

```
     mat_peptide     127. .546
                     /product="17kd antigen"
BASE COUNT      206 a    105 c    142 g    167 t
ORIGIN      Unreported.

  Query Match           13.2%;  Score 64.2;  DB 1;  Length 620;
  Best Local Similarity   49.6%;  Pred. No. 1.7e-07;
  Matches  193;  Conservative    0;  Mismatches  193;  Indels    3;  Gaps    1;

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
           ||||     ||| |     ||| | || ||  ||||   |   |||    |  || |||
Db     154 CAAGGTACAGGAACACTTCTTGGCGGTGCTGGCGGCGCATTACTTGGTTCTCAATTCGGT 213

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           || || |  || |   |||     |    ||  |||||  |  | || | | |  |||||
Db     214 AAGGGCAAAGGACAGCTTGTTGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGT 273

Qy     202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAA 261
             ||||||||  |   |||||| | ||||||| || |     | | |  | |   |
Db     274 GGACAAATCGGTGCAGGTATGGATGAACAGGATAGAAGACTTGCAGAGCTTACCTCACAG 333

Qy     262 AAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGT 321
           |  |    || |     ||   |  || |  ||||   |||     | |||    |||
Db     334 AGAGCTTTAGAAAC---AGCTCCTAGTGGTAGTAACGTAGAATGGCGTAATCCGGATAAC 390

Qy     322 GTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGT 381
           |   |    |          |    | | | | |   | |  |       ||||||||
Db     391 GGCAATTACGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGC 450

Qy     382 CGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCA 441
           || || |     ||||  |   |||| |  ||| |  | ||| |    |||||   ||||
Db     451 CGTGAGTACACTCAAACAGTTGTGATAGGCGGGAAACAACAAAAAGCATACGGTGATGCA 510

Qy     442 TGCCGGCAACCGGATGGTCGTTGGCAAGT 470
           ||||| ||||| || |  |  ||||||||
Db     511 TGCCGCCAACCTGACGAACAATGGCAAGT 539


RESULT 12
MRI269516
LOCUS       MRI269516               434 bp    DNA     linear   BCT 27-JAN-2001
DEFINITION  Male-killing Rickettsia from Adalia decempuntata partial 17kDA
            antigen gene.
ACCESSION   AJ269516
VERSION     AJ269516.1  GI:8920292
KEYWORDS    17 kDa antigen.
SOURCE      male-killing Rickettsia from Adalia decempunctata.
  ORGANISM  male-killing Rickettsia from Adalia decempunctata
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   von der Schulenburg,J.H., Habig,M., Sloggett,J.J., Webberley,K.M.,
            Bertrand,D., Hurst,G.D. and Majerus,M.E.
  TITLE     Incidence of male-killing Rickettsia spp. (alpha-proteobacteria) in
            the ten-spot ladybird beetle Adalia decempunctata L. (Coleoptera:
            Coccinellidae)
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 270-277 (2001)
  MEDLINE   20575219
   PUBMED   11133455
REFERENCE   2  (bases 1 to 434)
  AUTHORS   Schulenburg,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) Schulenburg H., Department of Evolutionary
            Biology, Institut fuer Spezielle Zoologie, Huefferstr. 1, 48149
            Muenster, GERMANY
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="male-killing Rickettsia from Adalia
                     decempunctata"
                     /specific_host="Adalia decempunctata"
                     /db_xref="taxon:120393"
                     /note="isolated from 10spot ladybird"
     CDS             <1. .>434
                     /codon_start=1
                     /transl_table=11
                     /product="17 kDA antigen"
                     /protein_id="CAB96381.1"
                     /db_xref="GI:8920293"
                     /db_xref="SPTREMBL:Q9K4W8"
                     /translation="ALATSMLQACTGPGGMNKQGTGTLLGGAGGALLGSQFGKGKGQL
                     VGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALEAAPSGSNVEWRNPDNGNHGY
                     VTPNKTYRNSTGQYCREYTQTVVIGGKQQKSYGNACRQPDEQ"
     primer_bind     1. .20
                     /note="PCR primer R1"
     primer_bind     415. .434
                     /note="PCR primer R2"
BASE COUNT      133 a     87 c    113 g    101 t
ORIGIN

  Query Match           12.8%;  Score 62.2;  DB 1;  Length 434;
  Best Local Similarity   50.1%;  Pred. No. 6.2e-07;
  Matches  182;  Conservative    0;  Mismatches  178;  Indels    3;  Gaps    1;

Qy     102 TGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGC 161
           ||| | ||  |  || | |   |  |||    |  || || |||||||  || | | |||
Db      75 TGGCGGTGCCGGCGGTGCATTACTTGGTTCTCAATTCGGCAAAGGTAAAGGGCAACTTGT 134

Qy     162 AATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGAT 221
           |    ||  |||||  |  | || | | |  |||||    ||||||||  |   ||
Db     135 CGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTAT 194

Qy     222 GGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGT 281
           |||| |||||||||| || |       | |  |  |   |||  |    || ||    ||
Db     195 GGATGAGCAGGATAGAAGACTTGCTGAACTCACATCTCAAAGAGCTTTAGAAGC---AGC 251

Qy     282 GACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTA 341
             |  || |  ||||   |||     | |||    |||  |   | |  |
Db     252 TCCTAGTGGTAGTAACGTAGAATGGCGTAATCCCGATAACGGCAATCATGGCTACGTAAC 311

Qy     342 CCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGC 401
             |    | | | | |   | |  |       |||||||| || || |     ||||  |
Db     312 ACCTAATAAAACTTATAGAAATAGCACAGGGCAATATTGCCGTGAGTACACTCAAACAGT 371

Qy     402 GATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCG 461
             | || |  ||  |  | ||| |    ||||| | |||||||| ||||| || |  |
Db     372 TGTAATAGGCGGAAAACAACAAAAATCATACGGTAATGCATGCCGCCAACCTGACGAACA 431

Qy     462 TTG 464
           |||
Db     432 ATG 434


RESULT 13
MRI269517
LOCUS       MRI269517               434 bp    DNA     linear   BCT 27-JAN-2001
DEFINITION  Male-killing Rickettsia from Adalia bipunctata partial 17kDA
            antigen gene (United Kingdom: Cambridge).
ACCESSION   AJ269517
VERSION     AJ269517.1  GI:8920294
KEYWORDS    17 kDa antigen.
SOURCE      male-killing Rickettsia from Adalia bipunctata.
  ORGANISM  male-killing Rickettsia from Adalia bipunctata
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   von der Schulenburg,J.H., Habig,M., Sloggett,J.J., Webberley,K.M.,
            Bertrand,D., Hurst,G.D. and Majerus,M.E.
  TITLE     Incidence of male-killing Rickettsia spp. (alpha-proteobacteria) in
            the ten-spot ladybird beetle Adalia decempunctata L. (Coleoptera:
            Coccinellidae)
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 270-277 (2001)
  MEDLINE   20575219
   PUBMED   11133455
REFERENCE   2  (bases 1 to 434)
  AUTHORS   Schulenburg,H.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (03-APR-2000) Schulenburg H., Department of Evolutionary
            Biology, Institut fuer Spezielle Zoologie, Huefferstr. 1, 48149
            Muenster, GERMANY
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="male-killing Rickettsia from Adalia bipunctata"
                     /specific_host="Adalia bipunctata"
                     /db_xref="taxon:38028"
                     /country="United Kingdom: Cambridge"
                     /note="isolated from 2spot ladybird"
     CDS             <1. .>434
                     /codon_start=1
                     /transl_table=11
                     /product="17 kDA antigen"
                     /protein_id="CAB96382.1"
                     /db_xref="GI:8920295"
                     /db_xref="SPTREMBL:Q9K2N6"
                     /translation="ALATSMLQACTGPGGMNKQGTGTLLGGAGGALLGSQFGKGKGQL
                     VGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALEAAPSGSNVEWRNPDNGNHGY
                     VTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDEQ"
     primer_bind     1. .20
                     /note="PCR primer R1"
     primer_bind     415. .434
                     /note="PCR primer R2"
BASE COUNT      132 a     88 c    114 g    100 t
ORIGIN

  Query Match             12.8%;  Score 62.2;  DB 1;  Length 434;
  Best Local Similarity   50.1%;  Pred. No. 6.2e-07;
  Matches  182;  Conservative    0;  Mismatches  178;  Indels    3;  Gaps    1;

Qy    102 TGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGC 161
          ||| | ||  |  || |   |   |||     |  || || |||||||  || | | |||
Db     75 TGGCGGTGCCGGCGGTGCATTACTTGGTTCTCAATTCGGCAAAGGTAAAGGGCAACTTGT 134

Qy    162 AATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGAT 221
            |    ||  |||||  |  | || | | |  |||||    ||||||||  |   ||
Db    135 CGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTAT 194

Qy    222 GGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGT 281
          ||||  |||||||||  || |       | |  |  |   |||  |    || ||    ||
Db    195 GGATGAGCAGGATAGAAGACTTGCTGAACTCACCTCTCAAAGAGCTTTAGAAGC---AGC 251

Qy    282 GACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTA 341
            |  || |  ||||   |||     | |||    |||  |   | |  |
Db    252 TCCTAGTGGTAGTAACGTAGAATGGCGTAATCCCGATAACGGCAATCATGGCTACGTAAC 311

Qy    342 CCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGC 401
           |    | | | |   | |  |       |||||||| || || |     ||||  |
Db    312 ACCTAATAAAACTTATAGAAATAGCACAGGGCAATATTGCCGTGAGTACACTCAAACAGT 371

Qy    402 GATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCG 461
           | || |  ||  |  | ||| |    ||||| | |||||||| ||||| || |  |
Db    372 TGTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCATGCCGCCAACCTGACGAACA 431

Qy    462 TTG 464
            ||
Db    432 ATG 434


RESULT 14
MRI269518
LOCUS       MRI269518               434 bp    DNA     linear   BCT 27-JAN-2001
DEFINITION  Male-killing Rickettsia from Adalia bipunctata partial 17kDA
            antigen gene (Denmark: Ribe).
ACCESSION   AJ269518
VERSION     AJ269518.1  GI:8920296
KEYWORDS    17 kDa antigen.
SOURCE      male-killing Rickettsia from Adalia bipunctata.
  ORGANISM  male-killing Rickettsia from Adalia bipunctata
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   von der Schulenburg,J.H., Habig,M., Sloggett,J.J., Webberley,K.M.,
            Bertrand,D., Hurst,G.D. and Majerus,M.E.
  TITLE     Incidence of male-killing Rickettsia spp. (alpha-proteobacteria) in
            the ten-spot ladybird beetle Adalia decempunctata L. (Coleoptera:
            Coccinellidae)
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 270-277 (2001)
  MEDLINE   20575219
   PUBMED   11133455
REFERENCE   2  (bases 1 to 434)
  AUTHORS   Schulenburg,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) Schulenburg H., Department of Evolutionary
            Biology, Institut fuer Spezielle Zoologie, Huefferstr. 1, 48149
            Muenster, GERMANY
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="male-killing Rickettsia from Adalia bipunctata"
                     /specific_host="Adalia bipunctata"
                     /db_xref="taxon:38028"
                     /country="Denmark: Ribe"
                     /note="isolated from 2spot ladybird"
     CDS             <1. .>434
                     /codon_start=1
                     /transl_table=11
                     /product="17kDA antigen"
                     /protein_id="CAB96383.1"
                     /db_xref="GI:8920297"
                     /db_xref="SPTREMBL:Q9K2N6"
                     /translation="ALATSMLQACTGPGGMNKQGTGTLLGGAGGALLGSQFGKGKGQL
                     VGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALEAAPSGSNVEWRNPDNGNHGY
                     VTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDEQ"
     primer_bind     1. .20
                     /note="PCR primer R1"
     primer_bind     415. .434
                     /note="PCR primer R2"
BASE COUNT      133 a     87 c    114 g    100 t
ORIGIN

  Query Match             12.8%;  Score 62.2;  DB 1;  Length 434;
  Best Local Similarity   50.1%;  Pred. No. 6.2e-07;
  Matches  182;  Conservative    0;  Mismatches  178;  Indels    3;  Gaps    1;

Qy    102 TGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGC 161
          ||| | ||  |  || |   |   |||     |  || || |||||||  || | | |||
Db     75 TGGCGGTGCCGGCGGTGCATTACTTGGTTCTCAATTCGGCAAAGGTAAAGGGCAACTTGT 134

Qy    162 AATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGTTCTAAAATCGGTCAATCGAT 221
            |    ||  |||||  |  | || | | |  |||||    ||||||||  |   ||
Db    135 CGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGTGGACAAATCGGTGCAGGTAT 194

Qy    222 GGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGT 281
          ||||  |||||||||  || |       | |  |  |   |||  |    || ||    ||
Db    195 GGATGAGCAGGATAGAAGACTTGCTGAACTCACATCTCAAAGAGCTTTAGAAGC---AGC 251

Qy    282 GACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGTGTTGAGCCAGTGCGTACTTA 341
            |  || |  ||||   |||     | |||    |||  |   | |  |
Db    252 TCCTAGTGGTAGTAACGTAGAATGGCGTAATCCCGATAACGGCAATCATGGCTACGTAAC 311

Qy    342 CCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGTCGAGAATTTCAGCAAAAGGC 401
           |    | | | |   | |  |       |||||||| || || |     ||||  |
Db    312 ACCTAATAAAACTTATAGAAATAGCACAGGGCAATATTGCCGTGAGTACACTCAAACAGT 371

Qy    402 GATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCATGCCGGCAACCGGATGGTCG 461
           | || |  ||  |  | ||| |    ||||| | |||||||| ||||| || |  |
Db    372 TGTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCATGCCGCCAACCTGACGAACA 431

Qy    462 TTG 464
            ||
Db    432 ATG 434


RESULT 15
AF195118
```

```
LOCUS       AF195118                546 bp    DNA     linear   BCT 12-JUN-2001
DEFINITION  Rickettsia felis 17 kDa genus-common antigen gene, complete cds.
ACCESSION   AF195118
VERSION     AF195118.1  GI:11066087
KEYWORDS    .
SOURCE      Rickettsia felis.
  ORGANISM  Rickettsia felis
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Bouyer,D.H., Stenos,J., Crocquet-Valdes,P., Moron,C.G., Popov,V.L.,
            Zavala-Velazquez,J.E., Foil,L.D., Stothard,D.R., Azad,A.F. and
            Walker,D.H.
  TITLE     Rickettsia felis: molecular characterization of a new member of the
            spotted fever group
  JOURNAL   Int. J. Syst. Evol. Microbiol. 51 (Pt 2), 339-347 (2001)
  MEDLINE   21217364
   PUBMED   11321078
REFERENCE   2  (bases 1 to 546)
  AUTHORS   Bouyer,D.H., Stenos,J., Crocquet-Valdes,P.A., Foil,L.D. and
            Walker,D.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1999) Pathology, University of Texas Medical
            Branch at Galveston, 301 University Blvd., Galveston, TX
            77551-0609, USA
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Rickettsia felis"
                     /db_xref="taxon:42862"
     CDS             65. .544
                     /codon_start=1
                     /transl_table=11
                     /product="17 kDa genus-common antigen"
                     /protein_id="AAG28452.1"
                     /db_xref="GI:11066088"
                     /translation="MKLLSKIMIIALAASMLQACNGPGGMNKQGTGTLLGGAGGALLG
                     SQFGKGKGQLVGVGVGALLGAVLGGQIGAGMDEQDRRLAELTSQRALEATPSGTSVEW
                     RNPDNGNHGYVTPNKTYRNSTGQYCREYTQTVVIGGKQQKAYGNACRQPDGLWQVVN"
BASE COUNT      183 a     96 c    126 g    141 t
ORIGIN
```

```
  Query Match            12.2%;  Score 59.4;  DB 1;  Length 546;
  Best Local Similarity  48.8%;  Pred. No. 4.2e-06;
  Matches  190;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    1;

Qy      82 CAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTGGCCAGCTGTTTGGT 141
           ||||     ||| |      ||| | ||  |  || |   |   |||    |  || ||
Db     149 CAAGGTACGGGAACACTTCTTGGCGGTGCCGGCGGTGCATTACTTGGTTCTCAATTCGGC 208

Qy     142 AAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGGGTGGATTAATTGGT 201
           || || |  || |   |||     |    ||  |||||  |  | || | | |  |||||
Db     209 AAGGGCAAAGGACAGCTTGTCGGAGTAGGTGTAGGTGCATTACTTGGAGCAGTTCTTGGT 268

Qy     202 TCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAAACCAGAGTTTGGAA 261
               |||| |||  |   |||||| |||||||||| || |       | |  | ||   ||
Db     269 GGACAAATAGGTGCAGGTATGGATGAGCAGGATAGAAGACTTGCTGAACTCACTTCACAA 328

Qy     262 AAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAGGCAATAGTTATAGT 321
           |  |    || ||    |  |   |  |   ||    |||     | |||    |||
Db     329 AGAGCTTTAGAAGCAACACCTA---GCGGCACTAGCGTAGAATGGCGTAATCCGGATAAC 385

Qy     322 GTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTCGCCAGCAATATTGT 381
           | | | |  |          |    | | | | |   | |  |       ||||||||
Db     386 GGTAATCATGGTTACGTAACACCTAATAAAACTTATAGAAATAGCACTGGTCAATATTGC 445

Qy     382 CGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGATTTACGGCACTGCA 441
           || || |     ||||  |   | || |  ||  |  | ||| |    ||||| | ||||
Db     446 CGTGAGTACACTCAAACAGTTGTAATAGGCGGAAAACAACAAAAAGCATACGGTAATGCA 505

Qy     442 TGCCGGCAACCGGATGGTCGTTGGCAAGT 470
           ||||| ||||| || || |  ||||||||
Db     506 TGCCGCCAACCTGACGGACTATGGCAAGT 534
```

Search completed: February 22, 2003, 16:09:02
Job time : 1555.32 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:08:35 ; Search time 134.862 Seconds
(without alignments)
8065.404 Million cell updates/sec

Title: US-09-677-374-3
Perfect score: 483
Sequence: 1 atgcgtggttgcctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	483	100.0	483	22	AAF86247	DNA sequence of E.
	2	483	100.0	483	24	ABK52402	DNA encoding Esche
	3	483	100.0	768	22	AAF86248	DNA sequence of c1
	4	483	100.0	768	24	ABK52403	E. coli codon opti
	5	295	61.1	486	22	AAF86246	OspA 17kD antigen
	6	295	61.1	486	24	ABK52401	DNA encoding outer
	7	295	61.1	489	22	AAH79040	Piscirickettsia sa
c	8	118	24.4	118	22	AAF86254	PCR primer #4 used
c	9	118	24.4	118	24	ABK52409	Escherichia coli c
	10	108.4	22.4	110	22	AAF86252	PCR primer #2 used
	11	108.4	22.4	110	24	ABK52407	Escherichia coli c
c	12	102	21.1	102	22	AAF86255	PCR primer #5 used
c	13	102	21.1	102	24	ABK52410	Escherichia coli c
	14	92.4	19.1	94	22	AAF86253	PCR primer #3 used
	15	92.4	19.1	94	24	ABK52408	Escherichia coli c
	16	74	15.3	111	22	AAF86251	PCR primer #1 used
	17	74	15.3	111	24	ABK52406	Escherichia coli c
c	18	72.4	15.0	110	22	AAF86256	PCR primer #6 used
c	19	72.4	15.0	110	24	ABK52411	Escherichia coli c
c	20	39.4	8.2	2292	21	AAC45883	Arabidopsis thalia
	21	36.6	7.6	303	16	AAT18913	DNA encoding spide
	22	36.6	7.6	303	16	AAT18912	DNA encoding spide
c	23	36	7.5	1305	17	AAT31799	Erythrobacter long
	24	36	7.5	3251	23	ABL03846	Drosophila melanog
	25	36	7.5	7791	22	AAL37335	Human musculoskele
c	26	35.6	7.4	292	21	AAA45539	Human secreted exp
c	27	35.6	7.4	2481	20	AAX90448	Human secreted pro
c	28	35.6	7.4	2481	22	AAS59277	Human cDNA encodin
c	29	35.6	7.4	2481	24	ABA90946	Human polynucleoti
c	30	35.4	7.3	571	20	AAX61355	DNA encoding a hum
	31	35.4	7.3	1290	19	AAV36619	Nucleotide sequenc
c	32	35.4	7.3	27541	22	AAD17185	Streptomyces nours
c	33	35.4	7.3	125401	22	AAD17186	Streptomyces nours
c	34	35.2	7.3	1479	21	AAA51609	HIV synthetic Gag
c	35	35.2	7.3	1479	24	ABL39953	Synthetic Gag poly
c	36	35	7.2	77536	21	AAA14651	Nucleotide sequenc
	37	34.8	7.2	9407	23	AAS59606	Propionibacterium
	38	34.6	7.2	77536	21	AAA14651	Nucleotide sequenc
	39	34.2	7.1	6273	23	ABL15731	Drosophila melanog
c	40	34.2	7.1	10855	23	ABL15730	Drosophila melanog
c	41	34	7.0	2652	20	AAX23386	Human Delta7-stero
c	42	34	7.0	2652	20	AAX23387	Human Delta7-stero
c	43	33.8	7.0	810	22	AAH66598	C glutamicum codin
c	44	33.8	7.0	933	22	AAF68080	Corynebacterium gl
	45	33.8	7.0	1062	24	ABQ90125	M. capsulatus gene

ALIGNMENTS

RESULT 1
AAF86247
ID AAF86247 standard; DNA; 483 BP.
XX
AC AAF86247;
XX
DT 11-JUL-2001 (first entry)
XX
DE DNA sequence of E. coli optimised ospA gene 17E2.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; ds.
XX
OS Piscirickettsia salmonis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..483
FT /*tag= a
FT /partial
FT /product= "OspA"
FT /note= "Genus specific 17kDa antigen, the sequence does
FT not include a stop codon"
XX
PN CA2281913-A1.
XX
PD 17-MAR-2001.
XX
PF 17-SEP-1999; 99CA-2281913.
XX
PR 17-SEP-1999; 99CA-2281913.

```
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81127.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4C; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents DNA which has been optimised for expression in
CC   Escherichia coli to encode the P. salmonis OspA protein. An OspA protein
CC   with an N-terminal fusion partner is used in a vaccine to create an
CC   anti-OspA antibody response.
XX
SQ   Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

  Query Match             100.0%;  Score 483;  DB 22;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-136;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60

Qy      61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120

Qy     121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180

Qy     181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240

Qy     241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300

Qy     301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360

Qy     361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420

Qy     421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480

Qy     481 AAA 483
          |||
Db     481 AAA 483


RESULT 2

ABK52402
ID   ABK52402 standard; DNA; 483 BP.
XX
AC   ABK52402;
XX
DT   12-AUG-2002  (first entry)
XX
DE   DNA encoding Escherichia coli codon optimised OspA, 17e2.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds; 17e2.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..483
FT                   /*tag=  a
FT                   /product= "17e2"
FT                   /note= "Escherichia coli codon optimised OspA"
FT                   /partial
FT                   /note= "No stop codon given"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97868.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 16; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Escherichia coli codon optimised outer surface
CC   lipoprotein OspA (17e2) used in the creation of the vaccine described in
CC   the invention.
XX
SQ   Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

  Query Match             100.0%;  Score 483;  DB 24;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-136;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60

Qy      61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120

Qy    121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180

Qy    181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240

Qy    241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300

Qy    301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360

Qy    361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420

Qy    421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480

Qy    481 AAA 483
          |||
Db    481 AAA 483


RESULT 3
AAF86248
ID   AAF86248 standard; DNA; 768 BP.
XX
AC   AAF86248;
XX
DT   11-JUL-2001  (first entry)
XX
DE   DNA sequence of c17e2 ospA construct with N-terminal fusion partner.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; ds.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..768
FT                   /*tag=  a
FT                   /partial
FT                   /product= "C17E2 OspA with N-terminal fusion partner"
FT                   /note= "No stop codon is given"
FT   misc_feature    1..285
FT                   /*tag=  b
FT                   /note= "DNA encoding undefined N-terminal fusion partner"
FT   misc_feature    286..768
FT                   /*tag=  c
FT                   /note= "Optimised OspA construct c17E2"
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81128.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 4; Fig 5; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents P. salmonis OspA DNA termed C17E2 optimised for
CC   expression in Escherichia coli fused to DNA encoding an undefined
CC   N-terminal fusion partner. The protein encoded by this fusion construct
CC   is used in a vaccine to create an anti-OspA antibody response.
XX
SQ   Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

  Query Match             100.0%;  Score 483;  DB 22;  Length 768;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-136;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    286 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 345

Qy     61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 405

Qy    121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 465

Qy    181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    466 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 525

Qy    241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    526 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 585

Qy    301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    586 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 645

Qy    361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    646 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 705

Qy    421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    706 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 765

Qy    481 AAA 483
          |||
Db    766 AAA 768


RESULT 4
ABK52403
ID   ABK52403 standard; DNA; 768 BP.
XX
AC   ABK52403;
```

```
XX
DT   12-AUG-2002  (first entry)
XX
DE   E. coli codon optimised OspA, 17e2 with N-terminal fusion DNA.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds; 17e2.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..768
FT                   /*tag=  a
FT                   /product= "17e2"
FT                   /note= "Escherichia coli codon optimised OspA with
FT                         N-terminal fusion"
FT                   /partial
FT                   /note= "No stop codon given"
FT   sig_peptide     1..285
FT                   /*tag=  a
FT                   /note= "This sequence represents the N-terminal fusion"
FT   mat_peptide     286..768
FT                   /*tag=  c
FT                   /label= "mature_17e2"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97869.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 16; Fig 5; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Escherichia coli codon optimised outer surface
CC   lipoprotein OspA (17e2) with an N-terminal fusion used in the creation of
CC   the vaccine described in the invention.
XX
SQ   Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

  Query Match             100.0%;  Score 483;  DB 24;  Length 768;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-136;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     286 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 345

Qy      61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     346 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 405

Qy     121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     406 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 465

Qy     181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     466 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 525

Qy     241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     526 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 585

Qy     301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     586 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 645

Qy     361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     646 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 705

Qy     421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     706 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 765

Qy     481 AAA 483
           |||
Db     766 AAA 768


RESULT 5
AAF86246
ID   AAF86246 standard; DNA; 486 BP.
XX
AC   AAF86246;
XX
DT   11-JUL-2001  (first entry)
XX
DE   OspA 17kD antigen gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; ds.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag=  a
FT                   /partial
FT                   /product= "OspA"
FT                   /note= "Genus specific 17kDa antigen, the sequence does
FT                         not include a stop codon"
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
```

```
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81126.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis -
XX
PS   Disclosure; Fig 2B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents DNA encoding the P. salmonis OspA protein. An OspA
CC   protein with an N-terminal fusion partner is used in a vaccine to create
CC   an anti-OspA antibody response.
XX
SQ   Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

  Query Match            61.1%;  Score 295;  DB 22;  Length 486;
  Best Local Similarity  76.0%;  Pred. No. 3.7e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64
Db       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy      65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124
Db      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy     125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184
Db     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy     185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244
Db     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy     245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304
Db     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy     305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364
Db     308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy     365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424
Db     368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy     425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483
Db     428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 6
ABK52401
ID   ABK52401 standard; DNA; 486 BP.
XX
AC   ABK52401;
XX
DT   12-AUG-2002  (first entry)
XX
DE   DNA encoding outer surface lipoprotein OspA.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag= a
FT                   /product= "OspA"
FT                   /note= "Outer surface lipoprotein"
FT                   /partial= "No stop codon given"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97867.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 19; Fig 2; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Piscirickettsia salmonis outer surface
CC   lipoprotein, OspA, used in the creation of the vaccine described in the
CC   invention.
XX
SQ   Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

  Query Match            61.1%;  Score 295;  DB 24;  Length 486;
  Best Local Similarity  76.0%;  Pred. No. 3.7e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy       5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64
Db       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy      65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124
Db      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy     125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184
Db     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy     185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244
Db     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy     245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304
Db     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307
```

```
Qy    305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364
          | || || ||   ||| || || || || || ||||||||||||||| || || || || |
Db    308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy    365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424
          | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db    368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy    425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483
          | |||||||| || |||| || |||||||| || ||||| || ||| || ||||||
Db    428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 7
AAH79040
ID   AAH79040 standard; cDNA; 489 BP.
XX
AC   AAH79040;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Piscirickettsia salmonis polynucleotide P10.6.
XX
KW   Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW   septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW   ATCC VR-1361; ss.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..489
FT                   /*tag=  a
XX
PN   WO200168865-A2.
XX
PD   20-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-GB01055.
XX
PR   11-MAR-2000; 2000GB-0005838.
PR   01-JUL-2000; 2000GB-0016080.
PR   01-JUL-2000; 2000GB-0016082.
PR   29-JUL-2000; 2000GB-0018599.
XX
PA   (AQUA-) AQUA HEALTH EURO LTD.
XX
PI   Simard N,  Brouwers H,  Jones S,  Griffiths S,  Valenzuela P;
PI   Burzio L;
XX
DR   WPI; 2001-639050/73.
DR   P-PSDB; AAG78025.
XX
PT   New nucleic acids encoding an amino acid sequence homologous to the
PT   surface antigen present on Piscirickettsia salmonis are useful to
PT   protect fish against piscirickettsiosis  -
XX
PS   Claim 3; Fig 4; 25pp; English.
XX
CC   The invention relates to nucleic acid sequences and the encoded protein
CC   of a least part of the surface antigen present on Piscirickettsia
CC   salmonis for production of a vaccine with antibacterial activity to
CC   protect fish against P. salmonis which causes piscirickettsiosis, also
CC   known as salmonid rickettsial septicaemia.
XX
SQ   Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other;

  Query Match             61.1%;  Score 295;  DB 22;  Length 489;
  Best Local Similarity   76.0%;  Pred. No. 3.7e-79;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy      5 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 64
          | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||

Db      8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy     65 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 124
          ||||||| || || || ||||| || || || || || || ||||| ||||||||||| |
Db     68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy    125 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 184
          |||||||||| ||||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db    128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy    185 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 244
          | ||  | ||||| |||||||||||||    ||||| ||||||||||||| || || |
Db    188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy    245 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 304
          |||||  | ||||||| || ||||| || || || || ||||||||||||| || || |
Db    248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy    305 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 364
          | || || ||   ||| || || || || || ||||||||||||||| || || || || |
Db    308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy    365 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 424
          | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db    368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy    425 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 483
          | |||||||| || |||| || |||||||| || ||||| || ||| || ||||||
Db    428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 8
AAF86254/c
ID   AAF86254 standard; DNA; 118 BP.
XX
AC   AAF86254;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #4 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
```

```
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;

  Query Match             24.4%;  Score 118;  DB 22;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-26;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     208 GGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAA 267
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 GGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAA 59

Qy     268 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAAC 325
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAAC 1


RESULT 9
ABK52409/c
ID   ABK52409 standard; DNA; 118 BP.
XX
AC   ABK52409;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #4.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Example 3; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence represents an oligonucleotide used to construct an
CC   Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC   used in the creation of the vaccine described in the invention.
XX
SQ   Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;

  Query Match             24.4%;  Score 118;  DB 24;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-26;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     208 GGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAA 267
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 GGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGGAAAAAGTGAAA 59

Qy     268 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAAC 325
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58 GCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTACTCTGTGGAAC 1


RESULT 10
AAF86252
ID   AAF86252 standard; DNA; 110 BP.
XX
AC   AAF86252;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #2 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

  Query Match             22.4%;  Score 108.4;  DB 22;  Length 110;
  Best Local Similarity   99.1%;  Pred. No. 5.5e-23;
  Matches  109;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      53 TGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGG 112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGG 60

Qy     113 GCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATG 162
           ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      61 GCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGGCGATG 110


RESULT 11
ABK52407
ID   ABK52407 standard; DNA; 110 BP.
XX
AC   ABK52407;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #2.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Example 3; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence represents an oligonucleotide used to construct an
CC   Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC   used in the creation of the vaccine described in the invention.
XX
SQ   Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

  Query Match             22.4%;  Score 108.4;  DB 24;  Length 110;
  Best Local Similarity   99.1%;  Pred. No. 5.5e-23;
  Matches  109;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      53 TGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGG 112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGG 60

Qy     113 GCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATG 162
           ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      61 GCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGGCGATG 110


RESULT 12
AAF86255/c
ID   AAF86255 standard; DNA; 102 BP.
XX
AC   AAF86255;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #5 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

  Query Match             21.1%;  Score 102;  DB 22;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-21;
  Matches  102;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     334 ACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAG 393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102 ACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAG 43

Qy     394 AAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 435
           |||||||||||||||||||||||||||||||||||||||||||
Db      42 AAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 1


RESULT 13
ABK52410/c
ID   ABK52410 standard; DNA; 102 BP.
```

```
XX
AC   ABK52410;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #5.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Example 3; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence represents an oligonucleotide used to construct an
CC   Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC   used in the creation of the vaccine described in the invention.
XX
SQ   Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

  Query Match             21.1%;  Score 102;  DB 24;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-21;
  Matches  102;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    334 ACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAG 393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    102 ACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAG 43

Qy    394 AAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 435
          |||||||||||||||||||||||||||||||||||||||||||
Db     42 AAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 1


RESULT 14
AAF86253
ID   AAF86253 standard; DNA; 94 BP.
XX
AC   AAF86253;
XX
DT   11-JUL-2001  (first entry)
XX
DE   PCR primer #3 used in cloning an optimisation of OspA gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; fusion construct; PCR primer; ss.
XX
OS   Piscirickettsia salmonis.
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents a PCR primer used in the cloning and optimisation of
CC   the P. salmonis OspA gene. The OspA gene is used in the method of the
CC   invention.
XX
SQ   Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

  Query Match             19.1%;  Score 92.4;  DB 22;  Length 94;
  Best Local Similarity   98.9%;  Pred. No. 3.6e-18;
  Matches   93;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    139 AAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 198
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAGGCTCTGGTCGTGTGGCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 60

Qy    199 TCTAAAATCGGTCAGAGCATGGACCAGCAGGATA 232
          |||||||||||||||||||||||||||||||||
Db     61 TCTAAAATCGGTCAGAGCATGGACCAGCAGGATA 94


RESULT 15
ABK52408
ID   ABK52408 standard; DNA; 94 BP.
XX
AC   ABK52408;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Escherichia coli codon optimised OspA 17e2, oligonucleotide #3.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; 17e2; ss.
XX
OS   Synthetic.
XX
```

PN CA2339327-A1.
XX
PD 15-MAR-2002.
XX
PF 19-MAR-2001; 2001CA-2339327.
XX
PR 15-SEP-2000; 2000US-0677374.
XX
PA (THOR/) THORNTON J C.
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZY/) KUZYK M A.
XX
PI Thornton JC, Kay WW, Burian J, Kuzyk MA;
XX
DR WPI; 2002-455221/49.
XX
PT Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT administration of an outer surface lipoprotein (OspA) of a bacterial
PT strain, as a vaccine -
XX
PS Example 3; Fig 4; 55pp; English.
XX
CC The invention describes a method of protecting a poikilothermic fish
CC against infection by the bacterial pathogen Piscirickettsia Salmonis
CC comprising administering either intraperitoneally, by immersion or
CC orally, an immunogenic amount of principal antigen, the OspA (outer
CC surface lipoprotein), its variants, non-lipidated form or antigenic
CC peptides derived or synthesized with or without an adjuvant. The new
CC method is used to provide an outer surface lipoprotein (OspA) of
CC bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC in fin-fish against Rickettsial septicaemia and other related
CC Rickettisial diseases caused by either a virus, bacteria or parasite.
CC This sequence represents an oligonucleotide used to construct an
CC Escherichia coli codon optimised outer surface lipoprotein OspA (17e2)
CC used in the creation of the vaccine described in the invention.
XX
SQ Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

Query Match 19.1%; Score 92.4; DB 24; Length 94;
Best Local Similarity 98.9%; Pred. No. 3.6e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    139 AAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 198
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 AAAGGCTCTGGTCGTGTGGCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 60

Qy    199 TCTAAAATCGGTCAGAGCATGGACCAGCAGGATA 232
          |||||||||||||||||||||||||||||||||
Db     61 TCTAAAATCGGTCAGAGCATGGACCAGCAGGATA 94
```

Search completed: February 22, 2003, 14:44:57
Job time : 141.862 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:26:45 ; Search time 957.936 Seconds
(without alignments)
8165.910 Million cell updates/sec

Title: US-09-677-374-3
Perfect score: 483
Sequence: 1 atgcgtggttgcctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	8.7	238	12	BG240789	BG240789 OV1_38_C0
2	42	8.7	402	12	BG356895	BG356895 OV2_11_D0
3	42	8.7	419	10	BE355894	BE355894 DG1_11_E1
4	42	8.7	428	12	BF176742	BF176742 EM1_4_F01
5	42	8.7	446	12	BG052290	BG052290 RHIZ2_12_
6	42	8.7	500	12	BF586874	BF586874 FM1_31_E0
7	42	8.7	507	12	BG053591	BG053591 RHIZ2_11_
8	42	8.7	531	12	BG053014	BG053014 RHIZ2_16_
9	42	8.7	537	12	BG713844	BG713844 EM1_20_C0
10	42	8.7	561	12	BF587805	BF587805 FM1_40_G0
11	42	8.7	577	12	BG673858	BG673858 EM1_20_C0
12	42	8.7	582	10	BE356763	BE356763 DG1_12_B1
13	42	8.7	591	10	BE356800	BE356800 DG1_12_E1
14	42	8.7	597	12	BG102589	BG102589 RHIZ2_34_
15	42	8.7	598	10	AW672446	AW672446 LG1_359_D
16	42	8.7	600	10	BE360868	BE360868 DG1_67_A0
17	42	8.7	620	12	BF481524	BF481524 FM1_21_B0
18	42	8.7	649	10	BE355895	BE355895 DG1_11_E1
19	40.4	8.4	548	13	BM328366	BM328366 PIC1_28_D
c 20	40.2	8.3	508	14	BQ242646	BQ242646 TaE15025B
c 21	40	8.3	513	10	AV430609	AV430609 AV430609
c 22	40	8.3	768	17	AQ690676	AQ690676 nbxb0082B
23	38	7.9	910	17	CNS006ON	AL065629 Drosophil
c 24	38	7.9	1101	17	CNS0100Y	AL098380 Drosophil
c 25	37.8	7.8	595	13	BI607335	BI607335 RH74382.5
c 26	37.8	7.8	826	13	BI522679	BI522679 603175711
27	37.8	7.8	925	17	CNS0091P	AL053013 Drosophil
c 28	37.6	7.8	1344	12	BE964680	BE964680 601658216
29	37.4	7.7	350	12	BG052289	BG052289 RHIZ2_12_
c 30	37.4	7.7	1101	17	CNS017SY	AL108460 Drosophil
c 31	37.2	7.7	399	14	BQ753608	BQ753608 EBan01_SQ
c 32	37.2	7.7	444	14	BQ753422	BQ753422 EBan01_SQ
c 33	37.2	7.7	514	12	BG320949	BG320949 Zm04_02f0
c 34	37	7.7	908	12	BF253465	BF253465 HVSMEf000
35	36.8	7.6	446	9	AU182544	AU182544 AU182544
c 36	36.8	7.6	1101	9	AL513871	AL513871 AL513871
37	36.4	7.5	604	14	BM936458	BM936458 UI-M-BH2.
c 38	36.4	7.5	1043	12	BG104511	BG104511 602311153
c 39	36.2	7.5	595	17	BH478914	BH478914 BOGMR38TF
40	36.2	7.5	647	13	BJ003370	BJ003370 BJ003370
41	36.2	7.5	663	13	BJ003303	BJ003303 BJ003303
c 42	36.2	7.5	789	9	AU125061	AU125061 AU125061
c 43	36	7.5	1058	12	BG295311	BG295311 602392604
44	35.8	7.4	682	10	BE593319	BE593319 WS1_100_C
c 45	35.6	7.4	432	14	BM821964	BM821964 K-EST0091

ALIGNMENTS

RESULT 1
BG240789
LOCUS BG240789 238 bp mRNA linear EST 15-FEB-2001
DEFINITION OV1_38_C06.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG240789
VERSION BG240789.1 GI:12775862
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 238)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTMix

```
            High quality sequence start: 3
            High quality sequence stop: 230
            POLYA=No.
FEATURES                Location/Qualifiers
     source             1. .238
                        /organism="Sorghum bicolor"
                        /db_xref="taxon:4558"
                        /clone_lib="Ovary 1 (OV1)"
                        /note="Organ: Mix of ovaries of varying immature stages
                        from 8-week-old plants; Vector: pBluescript II from Lambda
                        Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                        from poly-A RNA in the cloning vector lambda ZAP II.
                        Clones to be sequenced were prepared by mass excision."
BASE COUNT       63 a     54 c     70 g     51 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 238;
  Best Local Similarity   59.0%;  Pred. No. 0.027;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      79 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 138

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | |||  | || |   ||   |||||||   | || |||||
Db     139 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 198

Qy     180 TC 181
           |
Db     199 CC 200


RESULT 2
BG356895
LOCUS       BG356895                402 bp    mRNA    linear   EST 06-MAR-2001
DEFINITION  OV2_11_D01.g1_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG356895
VERSION     BG356895.1  GI:13238881
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 47
            High quality sequence stop: 391
            POLYA=No.
FEATURES                Location/Qualifiers
     source             1. .402
                        /organism="Sorghum bicolor"
                        /db_xref="taxon:4558"
                        /clone_lib="Ovary 2 (OV2)"
                        /note="Organ: Mix of ovaries of varying immature stages
                        from 8-week-old plants; Vector: pBluescript II from Lambda
                        Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                        from poly-A RNA in the cloning vector lambda ZAP II.
                        Clones to be sequenced were prepared by mass excision."
BASE COUNT      107 a     82 c    117 g     96 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 402;
  Best Local Similarity   59.0%;  Pred. No. 0.035;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     140 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 199

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | |||  | || |   ||   |||||||   | || |||||
Db     200 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 259

Qy     180 TC 181
           |
Db     260 CC 261


RESULT 3
BE355894
LOCUS       BE355894                419 bp    mRNA    linear   EST 20-JUL-2000
DEFINITION  DG1_11_E11.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355894
VERSION     BE355894.1  GI:9297451
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 16
            High quality sequence stop: 336
            POLYA=No.
FEATURES                Location/Qualifiers
     source             1. .419
                        /organism="Sorghum bicolor"
                        /db_xref="taxon:4558"
                        /clone_lib="Dark Grown 1 (DG1)"
                        /note="Organ: 5-day-old dark-grown seedlings; Vector:
                        Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                        made from poly-A RNA in the cloning vector lambda ZAP II.
                        Clones to be sequenced were prepared by mass excision."
BASE COUNT      102 a     82 c    116 g    119 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 10;  Length 419;
  Best Local Similarity   59.0%;  Pred. No. 0.036;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      78 CGGCGAGAAGCTGACCGACTAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 137

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | |||  | || |   ||   |||||||   | || |||||
```

```
Db    138 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 197

Qy    180 TC 181
          ||
Db    198 CC 199


RESULT 4
BF176742
LOCUS       BF176742                428 bp    mRNA    linear   EST 31-OCT-2000
DEFINITION  EM1_4_F01.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BF176742
VERSION     BF176742.1  GI:11064652
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 32
            High quality sequence stop: 427
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      103 a     80 c    119 g    126 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 428;
  Best Local Similarity   59.0%;  Pred. No. 0.036;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     58 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 117

Qy    120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
          | ||||||| |   | | |||  | || |   ||   ||||||||   | || ||||
Db    118 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 177

Qy    180 TC 181
          ||
Db    178 CC 179


RESULT 5
BG052290
LOCUS       BG052290                446 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_12_G04.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG052290
VERSION     BG052290.1  GI:12506830
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 11
            High quality sequence stop: 440
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      101 a     81 c    124 g    140 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 446;
  Best Local Similarity   59.0%;  Pred. No. 0.037;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     51 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 110

Qy    120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
          | ||||||| |   | | |||  | || |   ||   ||||||||   | || ||||
Db    111 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 170

Qy    180 TC 181
          ||
Db    171 CC 172


RESULT 6
BF586874
LOCUS       BF586874                500 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  FM1_31_E06.g1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
            propinquum cDNA, mRNA sequence.
ACCESSION   BF586874
VERSION     BF586874.1  GI:11679198
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: floral-induced meristems
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
```

```
          The University of Georgia, Department of Plant Biology
          Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
          Tel: 706 542 1860
          Fax: 706 583 0210
          Email: mmpratt@uga.edu
          Sequences have been trimmed to exclude PolyA, vector and regions
          below Phred quality 16. The threshold for highest quality sequence
          is 20.
          Seq primer: PolyTMix
          High quality sequence start: 8
          High quality sequence stop: 495
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Floral-Induced Meristem 1 (FM1)"
                     /note="Organ: Floral-induced meristems; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; mature plants were placed in a growth chamber for
                     15 days with 16 hr darkness  and 8 hr light (flowering is
                     induced by short-day conditions); 16 days after being
                     returned  to the greenhouse under natural long days during
                     late April/early May, meristems were  harvested The
                     library was made from poly-A RNA in the cloning vector
                     lambda ZAP II. Clones to be sequenced were prepared by
                     mass excision."
BASE COUNT      111 a     94 c    141 g    154 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 500;
  Best Local Similarity   59.0%;  Pred. No. 0.039;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      89 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 148

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     149 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 208

Qy     180 TC 181
           |
Db     209 CC 210


RESULT 7
BG053591
LOCUS       BG053591                507 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_11_C03.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG053591
VERSION     BG053591.1  GI:12509440
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: T7
            High quality sequence start: 95
            High quality sequence stop: 507
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      109 a     92 c    142 g    164 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 507;
  Best Local Similarity   59.0%;  Pred. No. 0.04;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      54 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 113

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     114 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 173

Qy     180 TC 181
           |
Db     174 CC 175


RESULT 8
BG053014
LOCUS       BG053014                531 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_16_D04.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG053014
VERSION     BG053014.1  GI:12508270
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 42
            High quality sequence stop: 527
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      122 a    101 c    152 g    154 t      2 others
ORIGIN
```

```
  Query Match             8.7%;  Score 42;  DB 12;  Length 531;
  Best Local Similarity   59.0%;  Pred. No. 0.04;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     121 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 180

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     181 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 240

Qy     180 TC 181
           |
Db     241 CC 242


RESULT 9
BG713844
LOCUS       BG713844                537 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  EM1_20_C08.g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG713844
VERSION     BG713844.1  GI:14007794
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 62
            High quality sequence stop: 536
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      118 a     98 c    152 g    169 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 537;
  Best Local Similarity   59.0%;  Pred. No. 0.041;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      92 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 151

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     152 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 211

Qy     180 TC 181
           |
Db     212 CC 213


RESULT 10
BF587805
LOCUS       BF587805                561 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  FM1_40_G05.g1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
            propinquum cDNA, mRNA sequence.
ACCESSION   BF587805
VERSION     BF587805.1  GI:11680115
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: floral-induced meristems
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 81
            High quality sequence stop: 561
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Floral-Induced Meristem 1 (FM1)"
                     /note="Organ: Floral-induced meristems; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; mature plants were placed in a growth chamber for
                     15 days with 16 hr darkness  and 8 hr light (flowering is
                     induced by short-day conditions); 16 days after being
                     returned  to the greenhouse under natural long days during
                     late April/early May, meristems were  harvested The
                     library was made from poly-A RNA in the cloning vector
                     lambda ZAP II. Clones to be sequenced were prepared by
                     mass excision."
BASE COUNT      140 a    111 c    160 g    150 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 561;
  Best Local Similarity   59.0%;  Pred. No. 0.042;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     202 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 261

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     262 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 321

Qy     180 TC 181
           |
Db     322 CC 323


RESULT 11
```

```
BG673858
LOCUS       BG673858                577 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  EM1_20_C08.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG673858
VERSION     BG673858.1  GI:13919511
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 57
            High quality sequence stop: 572
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      129 a    106 c    166 g    176 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 577;
  Best Local Similarity   59.0%;  Pred. No. 0.042;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     121 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 180

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | ||| | || |   ||   |||||||   | || |||||
Db     181 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 240

Qy     180 TC 181
             |
Db     241 CC 242


RESULT 12
BE356763
LOCUS       BE356763                582 bp    mRNA    linear   EST 20-JUL-2000
DEFINITION  DG1_12_B12.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE356763
VERSION     BE356763.1  GI:9298320
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 6
            High quality sequence stop: 553
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      145 a    116 c    166 g    155 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 10;  Length 582;
  Best Local Similarity   59.0%;  Pred. No. 0.042;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || | .  | | ||| || ||  | || ||
Db     213 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 272

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | ||| | || |   ||   |||||||   | || |||||
Db     273 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 332

Qy     180 TC 181
             |
Db     333 CC 334


RESULT 13
BE356800
LOCUS       BE356800                591 bp    mRNA    linear   EST 20-JUL-2000
DEFINITION  DG1_12_E10.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE356800
VERSION     BE356800.1  GI:9298357
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
```

```
          Sequences have been trimmed to exclude PolyA, vector and regions
          below Phred quality 16. The threshold for highest quality sequence
          is 20.
          Seq primer: PolyTMix
          High quality sequence start: 10
          High quality sequence stop: 557
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      146 a    114 c    168 g    163 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 10;  Length 591;
  Best Local Similarity   59.0%;  Pred. No. 0.043;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     210 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 269

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | |||  | || |   ||   |||||||   | || |||||
Db     270 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 329

Qy     180 TC 181
           |
Db     330 CC 331


RESULT 14
BG102589
LOCUS       BG102589                597 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  RHIZ2_34_H05.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG102589
VERSION     BG102589.1  GI:12617422
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 59
            High quality sequence stop: 593
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      149 a    130 c    170 g    148 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 12;  Length 597;
  Best Local Similarity   59.0%;  Pred. No. 0.043;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     267 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 326

Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
            | ||||||| |   |  | |||  | || |   ||   |||||||   | || |||||
Db     327 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 386

Qy     180 TC 181
           |
Db     387 CC 388


RESULT 15
AW672446
LOCUS       AW672446                598 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  LG1_359_D09.g1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW672446
VERSION     AW672446.1  GI:7536365
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 71
            High quality sequence stop: 598
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      151 a    128 c    173 g    146 t
ORIGIN

  Query Match             8.7%;  Score 42;  DB 10;  Length 598;
  Best Local Similarity   59.0%;  Pred. No. 0.043;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      60 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 119
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     279 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 338
```

```
Qy     120 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 179
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     339 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 398

Qy     180 TC 181
           |
Db     399 CC 400
```

Search completed: February 22, 2003, 17:05:56
Job time : 1004.94 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:36:28 ; Search time 40.0415 Seconds
(without alignments)
3699.288 Million cell updates/sec

Title: US-09-677-374-3
Perfect score: 483
Sequence: 1 atgcgtggttgcctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	36.6	7.6	303	4	US-08-556-978B-80	Sequence 80, Appl
	2	36.6	7.6	303	4	US-08-556-978B-81	Sequence 81, Appl
c	3	33.6	7.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	4	33.6	7.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	5	33.4	6.9	962	4	US-08-765-907A-16	Sequence 16, Appl
c	6	33.4	6.9	1052	2	US-08-403-852D-10	Sequence 10, Appl
c	7	33.4	6.9	1052	3	US-08-510-646B-10	Sequence 10, Appl
c	8	33.4	6.9	1052	4	US-09-231-818-10	Sequence 10, Appl
c	9	33.4	6.9	2888	4	US-08-765-907A-1	Sequence 1, Appli
	10	33.2	6.9	23673	4	US-09-773-816-1	Sequence 1, Appli
	11	33.2	6.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	12	33.2	6.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	13	33	6.8	981	4	US-09-307-973A-3	Sequence 3, Appli
c	14	33	6.8	1295	1	US-08-332-747-1	Sequence 1, Appli
c	15	32.4	6.7	535	6	5187077-4	Patent No. 5187077
c	16	32.4	6.7	535	6	5427925-3	Patent No. 5427925
c	17	32.4	6.7	599	6	5187077-24	Patent No. 5187077
c	18	32.4	6.7	599	6	5427925-22	Patent No. 5427925
c	19	32.4	6.7	714	2	US-08-378-617A-25	Sequence 25, Appl
c	20	32.4	6.7	1437	6	5187077-16	Patent No. 5187077
c	21	32.4	6.7	1437	6	5427925-14	Patent No. 5427925
	22	32.4	6.7	1820	4	US-09-530-212A-1	Sequence 1, Appli
c	23	32	6.6	1869	4	US-09-372-668-3	Sequence 3, Appli
c	24	32	6.6	2365	1	US-08-363-208-1	Sequence 1, Appli
c	25	32	6.6	2365	4	US-09-137-478-1	Sequence 1, Appli
	26	32	6.6	3300	1	US-08-194-290-6	Sequence 6, Appli
	27	32	6.6	3300	2	US-08-614-377A-6	Sequence 6, Appli
	28	32	6.6	3300	4	US-09-142-648B-6	Sequence 6, Appli
	29	31.8	6.6	13987	2	US-08-804-227C-13	Sequence 13, Appl
	30	31.8	6.6	44377	2	US-08-804-227C-7	Sequence 7, Appli
	31	31.8	6.6	44377	2	US-08-804-198-1	Sequence 1, Appli
c	32	31.6	6.5	453	4	US-09-397-787-239	Sequence 239, App
c	33	31.6	6.5	5036	4	US-09-177-349-2	Sequence 2, Appli
c	34	31.2	6.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c	35	30.8	6.4	1843	1	US-07-918-023-1	Sequence 1, Appli
	36	30.8	6.4	2171	3	US-08-851-843A-100	Sequence 100, App
	37	30.8	6.4	2171	4	US-08-974-549A-266	Sequence 266, App
	38	30.8	6.4	2171	4	US-08-854-050-100	Sequence 100, App
	39	30.8	6.4	2171	4	US-09-430-323-100	Sequence 100, App
	40	30.8	6.4	2810	4	US-09-408-647A-1	Sequence 1, Appli
c	41	30.6	6.3	882	4	US-08-818-112-138	Sequence 138, App
c	42	30.6	6.3	882	4	US-08-818-111-133	Sequence 133, App
c	43	30.6	6.3	882	4	US-09-056-556-138	Sequence 138, App
c	44	30.6	6.3	882	4	US-09-072-596-133	Sequence 133, App
c	45	30.6	6.3	1689	1	US-08-442-884-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-556-978B-80
; Sequence 80, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-556-978B-80

Query Match 7.6%; Score 36.6; DB 4; Length 303;
Best Local Similarity 57.4%; Pred. No. 0.053;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 85 GTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGC 144

```
Db    138 GCTGCGGCAGCTGGTGGTGCTGGTCAGGGCGGTCTTGGCTCACAAGGGGCCGGTCAAGGC 197

Qy    145 TCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCT 199
          |||||   |   |    |||   |||||||| |  |  || ||       |||||
Db    198 GCTGGTGCAGCAGCAGCTGCCGCTGGCGGTGCAGGCCAAGGTGGATATGGTGGCT 252


RESULT 2
US-08-556-978B-81
; Sequence 81, Application US/08556978B
; Patent No. 6268169
;  GENERAL INFORMATION:
;    APPLICANT:  FAHNESTOCK, STEPHEN F.
;    TITLE OF INVENTION:  NOVEL RECOMBINANTLY PRODUCED
;    TITLE OF INVENTION:  SPIDER SILK ANALOGS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD FOR WINDOWS 95
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/556,978B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/077,600
;      FILING DATE:  JUNE 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9389-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  81:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-556-978B-81

  Query Match             7.6%;  Score 36.6;  DB 4;  Length 303;
  Best Local Similarity   57.4%;  Pred. No. 0.053;
  Matches   66;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy     85 GTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGC 144
          | ||  || ||    ||||| | |  ||||||| ||| |   ||   |  |||| |||||
Db     87 GCTGCGGCAGCTGGTGGTGCTGGTCAGGGCGGTCTTGGCTCACAAGGGGCCGGTCAAGGC 146

Qy    145 TCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCT 199
          |||||   |   |    |||   |||||||| |  |  || ||       |||||
Db    147 GCTGGTGCAGCAGCAGCTGCCGCTGGCGGTGCAGGCCAAGGTGGATATGGTGGCT 201


RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             7.0%;  Score 33.6;  DB 4;  Length 4403765;
  Best Local Similarity   51.3%;  Pred. No. 22;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     82 GAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAA 141
          || || |  || ||| |||||||| |  ||| | |  || | || | | |  ||||||
Db 2884704 GATGTGGTTGCCGCCGCCGGTGCCGCCGTGCGTGCCGTCGACGTCGAACGCTTCGGTGGT 2884645

Qy    142 GGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCT 201
          |  || || |  | |    ||   ||    |||      |||       || |  ||
Db 2884644 GAGTCGGGCCAGGCGCGCTTGTTGATACAGGGTTTCCAGCTGTTGCCGTTGTTCTGCAAA 2884585

Qy    202 AAAATCGGTCAGAGCATGGACCAGCAGGATAA 233
          |||| ||||  | ||   |   | |  ||| |
Db 2884584 AAAAGCGGTGGGCGCGATGTTGACCTCGATGA 2884553


RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             7.0%;  Score 33.6;  DB 4;  Length 4411529;
  Best Local Similarity   48.9%;  Pred. No. 22;
  Matches   90;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     39 CTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCAC 98
          | |||    |||| ||  | | |||     |  | || |    |   | ||||  |||
Db 2061470 CACTGCGGGCCTGCTGTTTGGTGCCGGCGGCGCCGGCGGATTCGGCTTCGGCGGTGCCGG 2061411

Qy     99 CGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTC 158
          || || | || |  ||| |      | |  |||| ||||  | |||  ||   | |   |
Db 2061410 GGGCGCCGGTGGGCTCGGCGGCAAAGCCGGGCTGATCGGCGACGGCGGTGACGGCGGCGC 2061351

Qy    159 GATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCAT 218
             | | ||||   | |    |||||||      ||| ||     ||||   | | |
Db 2061350 CGGAGGAAACGGCACCGGTGCCAAGGGCGGTGACGGCGGCGCTGGCGGCGGTGCCATCCT 2061291
```

```
Qy    219 GGAC 222
          || |
Db 2061290 GGTC 2061287


RESULT 5
US-08-765-907A-16/c
; Sequence 16, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
;  APPLICANT: BLANC, Veronique
;  APPLICANT:  THIBAUT, Denis
;  APPLICANT:  BAMAS-JACQUES, Nathalie
;  APPLICANT:  BLANCHE, Francis
;  APPLICANT:  COUZET, Joel
;  APPLICANT:  BARRIERE, Jean-Claude
;  APPLICANT:  DEBUSSCHE, Laurent
;  APPLICANT:  FAMECHON, Alain
;  APPLICANT:  PARIS, Jean-Marc
;  APPLICANT:  DUTRUC-ROSSET, Gilles
;  TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
;  TITLE OF INVENTION:  Mutasynthesis
;  FILE REFERENCE: Streptogramin genes
;  CURRENT APPLICATION NUMBER: US/08/765,907A
;  CURRENT FILING DATE:  1997-03-20
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 962
;   TYPE: DNA
;   ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-16

  Query Match             6.9%;  Score 33.4;  DB 4;  Length 962;
  Best Local Similarity   55.7%;  Pred. No. 0.82;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy     75 CCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTT 134
          | |||||| || || || || | || | |  |   |||  || || | |||||||| |
Db    243 CGGCCAGGCAGGTGTCGAGGTCCCCCGGGGCGACCTGGCAGGCGTGGGCGGCCAGCAGCC 184

Qy    135 CGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGT 189
           |||   ||| | ||  | ||       |    |||||  |     |||  ||||
Db    183 GGGTGTCGGCCCGGGCGGCGTCGGCGGTGATCCCGGCGCCGGTCAGCTGCCCGGT 129


RESULT 6
US-08-403-852D-10/c
; Sequence 10, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  84..962
;      OTHER INFORMATION:  /product= "Gene PapM"
US-08-403-852D-10

  Query Match             6.9%;  Score 33.4;  DB 2;  Length 1052;
  Best Local Similarity   55.7%;  Pred. No. 0.86;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy     75 CCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTT 134
          | |||||| || || || || | || | |  |   |||  || || | |||||||| |
Db    243 CGGCCAGGCAGGTGTCGAGGTCCCCCGGGGCGACCTGGCAGGCGTGGGCGGCCAGCAGCC 184

Qy    135 CGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGT 189
           |||   ||| | ||  | ||       |    |||||  |     |||  ||||
Db    183 GGGTGTCGGCCCGGGCGGCGTCGGCGGTGATCCCGGCGCCGGTCAGCTGCCCGGT 129


RESULT 7
US-08-510-646B-10/c
; Sequence 10, Application US/08510646B
; Patent No. 6077699
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
```

```
;       ZIP:  20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/510,646B
;       FILING DATE:  03-AUG-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/403,852
;       FILING DATE:  10-MAY-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/FR 93/00923
;       FILING DATE:  25-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 92/11441
;       FILING DATE:  25-SEP-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meyers, Kenneth J.
;       REGISTRATION NUMBER:  25,146
;       REFERENCE/DOCKET NUMBER:  03806.0054-01000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 408-4000
;       TELEFAX:  (202) 408-4400
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1052 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  S.pristinaespiralis
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  84..962
;       OTHER INFORMATION:  /product= "Gene PapM"
US-08-510-646B-10

  Query Match              6.9%;  Score 33.4;  DB 3;  Length 1052;
  Best Local Similarity   55.7%;  Pred. No. 0.86;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      75 CCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTT 134
           | |||||| || || || || | || | |  |   |||  || || | |||||||| |
Db     243 CGGCCAGGCAGGTGTCGAGGTCCCCCGGGGCGACCTGGCAGGCGTGGGCGGCCAGCAGCC 184

Qy     135 CGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGT 189
           |||   ||| | ||  | ||       |    |||||  |     |||  ||||
Db     183 GGGTGTCGGCCCGGGCGGCGTCGGCGGTGATCCCGGCGCCGGTCAGCTGCCCGGT 129


RESULT 8
US-09-231-818-10/c
; Sequence 10, Application US/09231818
; Patent No. 6171846
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/231,818
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852
;      FILING DATE:  10-MAY-1995
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1052 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  84..962
;      OTHER INFORMATION:  /product= "Gene PapM"
US-09-231-818-10

  Query Match              6.9%;  Score 33.4;  DB 4;  Length 1052;
  Best Local Similarity   55.7%;  Pred. No. 0.86;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      75 CCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTT 134
           | |||||| || || || || | || | |  |   |||  || || | |||||||| |
Db     243 CGGCCAGGCAGGTGTCGAGGTCCCCCGGGGCGACCTGGCAGGCGTGGGCGGCCAGCAGCC 184

Qy     135 CGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGT 189
           |||   ||| | ||  | ||       |    |||||  |     |||  ||||
Db     183 GGGTGTCGGCCCGGGCGGCGTCGGCGGTGATCCCGGCGCCGGTCAGCTGCCCGGT 129


RESULT 9
US-08-765-907A-1/c
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
;  APPLICANT: BLANC, Veronique
;  APPLICANT:  THIBAUT, Denis
;  APPLICANT:  BAMAS-JACQUES, Nathalie
;  APPLICANT:  BLANCHE, Francis
;  APPLICANT:  COUZET, Joel
;  APPLICANT:  BARRIERE, Jean-Claude
;  APPLICANT:  DEBUSSCHE, Laurent
```

```
;  APPLICANT:  FAMECHON, Alain
;  APPLICANT:  PARIS, Jean-Marc
;  APPLICANT:  DUTRUC-ROSSET, Gilles
;  TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
;  TITLE OF INVENTION:  Mutasynthesis
;  FILE REFERENCE: Streptogramin genes
;  CURRENT APPLICATION NUMBER: US/08/765,907A
;  CURRENT FILING DATE:  1997-03-20
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2888
;   TYPE: DNA
;   ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

  Query Match             6.9%;  Score 33.4;  DB 4;  Length 2888;
  Best Local Similarity   55.7%;  Pred. No. 1.3;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      75 CCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTT 134
           | |||||| || || || || | || | |   |   |||  || || | |||||||| |
Db    2418 CGGCCAGGCAGGTGTCGAGGTCCCCCGGGGCGACCTGGCAGGCGTGGGCGGCCAGCAGCC 2359

Qy     135 CGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGT 189
           |||   ||| | ||  | ||       |    |||||  |     |||  ||||
Db    2358 GGGTGTCGGCCCGGGCGGCGTCGGCGGTGATCCCGGCGCCGGTCAGCTGCCCGGT 2304


RESULT 10
US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
;  APPLICANT: Stanford University
;  APPLICANT:  Khosla, Chaitan
;  TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
;  TITLE OF INVENTION:  ANTAGONISTS
;  FILE REFERENCE: 28600-20210.00
;  CURRENT APPLICATION NUMBER: US/09/773,816
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/243,458
;  PRIOR FILING DATE: 2000-10-25
;  PRIOR APPLICATION NUMBER: US 60/179,305
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 23673
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(23623)
;   OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

  Query Match             6.9%;  Score 33.2;  DB 4;  Length 23673;
  Best Local Similarity   52.1%;  Pred. No. 3.8;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      55 GGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGC 114
           || | || | |   |   |  |   ||||     |  |    |  ||  || |  || ||
Db    8266 GGGTTCGGCGACGATGACGTCGCACAGGGTCGCGAAGGACGCCGCGACGCCGCCGTAGGT 8325

Qy     115 GGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGT 174
           || || |  | |||||     ||| ||  | | || |  ||   | || ||   |||||
Db    8326 GGGGTCGGTGACCAGCGAGATGGTCAACACGCCGGCCTCGTCCAGTTGCCCGAGGGCGGC 8385

Qy     175 GCGGTTCTGGGCGGTCTGATTG 196
           || | ||| |||  ||||  ||
Db    8386 GCTGGTCTTGGCCATCTGCATG 8407


RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             6.9%;  Score 33.2;  DB 4;  Length 4403765;
  Best Local Similarity   50.6%;  Pred. No. 28;
  Matches   80;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      59 GCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTG 118
           |||||  |  | | | | | | |    | | ||||  || ||   |  |  |||  |
Db 835437 GCGCCGCGTCCATAACCAGTCCGCTGCTCGACGCGATCAACGCGCCCTTCCTGGCGGCGT 835496

Qy     119 TTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGG 178
           | |   ||| |||| ||||||| ||| | |   | | | ||  | ||   | ||  | |
Db 835497 TGGGGCGCCCGCTGATCGGTAACGGCGCCGACGGGGCGCCGGGGACCGGGGCCGCCGGCG 835556

Qy     179 TTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGC 216
                 |||||  || |   |   ||   |||   | ||
Db 835557 GGGCCGGCGGATTGTTGTTCGGCAACGGCGGCGCGGGC 835594


RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             6.9%;  Score 33.2;  DB 4;  Length 4411529;
  Best Local Similarity   50.6%;  Pred. No. 28;
  Matches   80;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      59 GCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTG 118
           |||||  |  | | | | | | |    | | ||||  || ||   |  |  |||  |
```

```
Db 833259 GCGCCGCGTCCATAACCAGTCCGCTGCTCGACGCGATCAACGCGCCCTTCCTGGCGGCGT 833318

Qy    119 TTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGG 178
          | |   ||| |||| ||||||| ||| | |   | | | ||  | ||   | ||  | |
Db 833319 TGGGGCGCCCGCTGATCGGTAACGGCGCCGACGGGGCGCCGGGGACCGGGGCCGCCGGCG 833378

Qy    179 TTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGC 216
                |||||  || |   |   ||   |||   | ||
Db 833379 GGGCCGGCGGATTGTTGTTCGGCAACGGCGGCGCGGGC 833416


RESULT 13
US-09-307-973A-3
; Sequence 3, Application US/09307973A
; Patent No. 6432686
;  GENERAL INFORMATION:
;    APPLICANT:  BULTHIUS, BEN
;    APPLICANT:  GATENBY, ANTHONY A.
;    APPLICANT:  TRIMBUR, DONALD E.
;    APPLICANT:  WHITED, GREGORY
;    TITLE OF INVENTION:  METHOD FOR THE PRODUCTION OF
;    TITLE OF INVENTION:  1,3-PROPANEDIOL BY RECOMBINANT
;    TITLE OF INVENTION:  ORGANISMS COMPRISING GENES FOR
;    TITLE OF INVENTION:  VITAMIN B12 TRANSPORT
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DUPONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;      ADDRESSEE:  GENENCOR INTERNATIONAL, INC.
;      STREET:  925 PAGE MILL ROAD
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 95
;      SOFTWARE:  MICROSOFT OFFICE 97
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/307,973A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/085,190
;      FILING DATE:  JUNE 30, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1245
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  981 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-307-973A-3

  Query Match             6.8%;  Score 33;  DB 4;  Length 981;
  Best Local Similarity   53.5%;  Pred. No. 1.1;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     37 ATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCC 96
          |||||  |     |   ||||||   |||  |   |||  | |||  | |   | |   || ||
Db    655 ATCTCGGCGCGGCAACTGGGTTTACCCCTGTGGTTCTGGCGCAATGTGCTGGTGGCAGCG 714

Qy     97 ACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTG 156
          |||||   | | || ||||    ||  |  | |  |   | ||  |||| |  | || ||
Db    715 ACCGGCTGGATGGTTGGCGTCAGTGTGGCGCTGGCGGGTGCTATCGGCTTTATTGGTCTG 774

Qy    157 TCGATGGCC 165
             |||  ||
Db    775 GTGATCCCC 783


RESULT 14
US-08-332-747-1/c
; Sequence 1, Application US/08332747
; Patent No. 5733724
;  GENERAL INFORMATION:
;    APPLICANT:  Bubert, Andreas
;    APPLICANT:  Goebel, Werner
;    APPLICANT:  Goetz, Monika
;    APPLICANT:  Ludwig, Albrecht
;    APPLICANT:  Schubert, Peter
;    APPLICANT:  Neumann, Siegfried
;    TITLE OF INVENTION:  Oligonucleotides for the detection of
;    TITLE OF INVENTION:  Enterobacteriaceae
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MILLEN, WHITE, ZELANO & BRANIGAN
;      STREET:  2200 CLARENDON BLVD., SUITE 1400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,747
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 43 37 295
;      FILING DATE:  02-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hamlet-King, Diana
;      REGISTRATION NUMBER:  33,302
;      REFERENCE/DOCKET NUMBER:  MERCK 1628
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-243-6333
;      TELEFAX:  703 243-6410
;      TELEX:  64191
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Salmonella sp.
US-08-332-747-1

  Query Match             6.8%;  Score 33;  DB 1;  Length 1295;
  Best Local Similarity   55.8%;  Pred. No. 1.2;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     86 TTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCT 145
          |||||||   |  ||| |||||  |||||||  |   |||| |     ||||    ||
```

```
Db    942 TTGGCGCTATCGGCGGCGCGGTGTTGGGCGGATTCCTCGGCAACACTATCGGCGGCGGTA 883

Qy    146 CTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 198
          | ||||||     |  | ||   |  || |||||    |||||  |    |||
Db    882 CGGGTCGTTCACTGGCGACCGCGGCAGGCGCGGTGGCTGGCGGCGTAGCGGGC 830


RESULT 15
5187077-4/c
;Patent No. 5187077
;    APPLICANT: GEARING, DAVID P.;GOUGH, NICHOLAS M.;HILTON,
;DOUGLAS J.;KING, JULIE A.;METCALF, DONALD;NICE, EDOUARD C.
;NICOLA, NICOS A.;SIMPSON, RICHARD J.;WILLSON, TRACY A.
;    TITLE OF INVENTION: LEUKEMIA INHIBITORY FACTOR
;    NUMBER OF SEQUENCES: 41
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/294,514
;      FILING DATE: 09-DEC-1988
;SEQ ID NO:4:
;      LENGTH: 535
5187077-4

  Query Match             6.7%;  Score 32.4;  DB 6;  Length 535;
  Best Local Similarity   47.9%;  Pred. No. 1.3;
  Matches   93;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     32 TCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCG 91
          || || ||   ||  ||  ||   | ||   |    || |   || |  || | |||  |
Db    482 TCTTTGTCAGAGTGGTCGGGGACAGGTGGCACATCCACGTGGCCCACACGGTACTTGTTG 423

Qy     92 CGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTC 151
          |    || |    |    | |  |  |  |||   ||     ||  ||   ||  ||  |
Db    422 CACAGACGGCAAAGCACATTGCTGAGGAGGCCCCTCATGACGTCTATAGTAGCATTGAGC 363

Qy    152 GTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTC 211
           ||    |  |||   |||| ||| ||||| || |  |||| |      | | || ||||
Db    362 TTGACCTGGAGGCTCACGGCAGTGGGGTTCAGGACCTTCTGGTCCCGGGTGATATTGGTC 303

Qy    212 AGAGCATGGACCAG 225
          || |    |  |||
Db    302 AGGGAGGCGCTCAG 289


Search completed: February 22, 2003, 18:48:18
Job time : 6109.04 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 16:18:29 ; Search time 46.9931 Seconds
(without alignments)
5772.590 Million cell updates/sec

Title: US-09-677-374-3
Perfect score: 483
Sequence: 1 atgcgtggttgcctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	7.5	7791	10	US-09-764-877-3700	Sequence 3700, Ap
c 2	35.6	7.4	2481	10	US-09-729-674-141	Sequence 141, App
3	34.8	7.2	426	9	US-10-125-815-6	Sequence 6, Appli
c 4	33.8	7.0	810	9	US-09-738-626-1633	Sequence 1633, Ap
5	33.6	7.0	1164	9	US-09-738-626-2633	Sequence 2633, Ap
6	33.2	6.9	1047	10	US-09-780-996-6	Sequence 6, Appli
7	33.2	6.9	1278	9	US-10-086-738A-1	Sequence 1, Appli
8	33.2	6.9	1792	10	US-09-822-830A-420	Sequence 420, App
c 9	32.8	6.8	900	9	US-09-712-363-88	Sequence 88, Appl
c 10	32	6.6	1869	9	US-10-115-195-3	Sequence 3, Appli
c 11	32	6.6	2364	9	US-10-057-609B-1	Sequence 1, Appli
c 12	32	6.6	2863	9	US-09-822-846-108	Sequence 108, App
13	32	6.6	3300	10	US-09-379-931-6	Sequence 6, Appli
c 14	31.8	6.6	891	9	US-09-738-626-2169	Sequence 2169, Ap
c 15	31.8	6.6	1350	10	US-09-963-690-1	Sequence 1, Appli
c 16	31.6	6.5	453	10	US-09-876-889-239	Sequence 239, App
c 17	31.6	6.5	5036	10	US-09-918-951-2	Sequence 2, Appli
c 18	31.4	6.5	3919	10	US-09-070-927A-681	Sequence 681, App
c 19	31.2	6.5	1204	10	US-09-822-849A-22	Sequence 22, Appl
20	31.2	6.5	2678	10	US-09-964-824A-252	Sequence 252, App
c 21	31.2	6.5	4397	10	US-09-728-952-43	Sequence 43, Appl
c 22	31.2	6.5	4397	10	US-09-728-952-98	Sequence 98, Appl
c 23	31	6.4	116	9	US-10-025-380-856	Sequence 856, App
c 24	31	6.4	116	10	US-09-922-217-856	Sequence 856, App
c 25	31	6.4	116	10	US-09-833-263-856	Sequence 856, App
26	31	6.4	147	10	US-09-974-300-3732	Sequence 3732, Ap
27	30.8	6.4	2171	9	US-09-843-676-100	Sequence 100, App
28	30.8	6.4	2171	9	US-09-766-253-100	Sequence 100, App
29	30.8	6.4	2171	9	US-09-438-486-100	Sequence 100, App
30	30.8	6.4	2171	9	US-10-053-758-100	Sequence 100, App
c 31	30.6	6.3	4030	9	US-09-990-046-1	Sequence 1, Appli
32	30.6	6.3	43058	10	US-09-954-456-292	Sequence 292, App
33	30.6	6.3	43058	10	US-09-954-456-529	Sequence 529, App
34	30.6	6.3	43058	10	US-09-880-107-3950	Sequence 3950, Ap
c 35	30.4	6.3	411	10	US-09-864-761-27332	Sequence 27332, A
c 36	30.4	6.3	492	10	US-09-864-761-10686	Sequence 10686, A
c 37	30.4	6.3	1317	9	US-09-712-363-105	Sequence 105, App
38	30.2	6.3	509	10	US-09-864-761-12054	Sequence 12054, A
39	30.2	6.3	1035	9	US-10-034-158-8	Sequence 8, Appli
40	30.2	6.3	1035	10	US-09-086-118-28	Sequence 28, Appl
41	30.2	6.3	1035	10	US-09-835-684-12	Sequence 12, Appl
42	30.2	6.3	1035	10	US-09-880-371-12	Sequence 12, Appl
43	30.2	6.3	1035	10	US-09-879-248-16	Sequence 16, Appl
44	30.2	6.3	1035	10	US-09-770-693-8	Sequence 8, Appli
45	30.2	6.3	1035	10	US-09-766-348-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-764-877-3700
; Sequence 3700, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3700
; LENGTH: 7791
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3700

Query Match 7.5%; Score 36; DB 10; Length 7791;
Best Local Similarity 51.2%; Pred. No. 0.35;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy      35 TTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGG 94
           ||  || |  ||  |    ||||   |||| ||    | |  |||    || ||  | ||
Db    1025 TTCCCTTTCCTTGGCCTTGGGGTGCGGCCCGGACGGGCGGGGGCCGAACAGGTGCGGGGG 1084

Qy      95 CCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTG 154
           ||  ||| ||||  |  || ||    || ||    |||| | |  | |||   ||  | |
Db    1085 CCCGCGGGGCGGGGGAAGGGGGCCGGGCTGGAGGCCTGTGCAGCGAGGGCGAGGGCGGAG 1144

Qy     155 TGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 198
           |  ||  |  |||||||  | |    |  || |||  | |
Db    1145 GGAGGAGCGGGCTCGGCGGGAGGTTGGGAGCAGGGCTGACAGCC 1188
```

RESULT 2
US-09-729-674-141/c
; Sequence 141, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth

```
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6055-64X
;  CURRENT APPLICATION NUMBER: US/09/729,674
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/539,330
;  PRIOR FILING DATE: 2000-03-30
;  NUMBER OF SEQ ID NOS: 283
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 141
;   LENGTH: 2481
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-729-674-141

  Query Match             7.4%;  Score 35.6;  DB 10;  Length 2481;
  Best Local Similarity   52.7%;  Pred. No. 0.27;
  Matches   77;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy      59 GCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTG 118
           || ||  ||| | | || | ||| |    ||  |||  |       ||| | || || |
Db     402 GCACCTGGAAGGTGACCCACAAGGTATAGAGCTGGGCGGCTTTCCTCGTTATAGGTGGAG 343

Qy     119 TTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGG 178
           |    | |||| ||| ||  |   |  |  |||  ||| ||||| ||| |   ||  |||
Db     342 TCTTGGCCCAGATGTCCGAGAGCCGAGCATGTCCGGTGACGATGTCCACCACAGGGCCGG 283

Qy     179 TTCTGGGCGGTCTGATTGGCTCTAAA 204
           |   ||     |||  | | ||  ||
Db     282 TCAGGGCGCAGCTGTATTGGTCACAA 257


RESULT 3
US-10-125-815-6
; Sequence 6, Application US/10125815
; Patent No. US20020173008A1
; GENERAL INFORMATION:
;  APPLICANT: Hu, Zhihao
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Santi, Daniel V.
;  TITLE OF INVENTION: RECOMBINANT POLYKETIDE SYNTHASE GENES
;  FILE REFERENCE: 300622004720
;  CURRENT APPLICATION NUMBER: US/10/125,815
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 09/679,279
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: 60/190,024
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/158,305
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: megAT6
US-10-125-815-6

  Query Match             7.2%;  Score 34.8;  DB 9;  Length 426;
  Best Local Similarity   47.7%;  Pred. No. 0.21;
  Matches  102;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy      77 GCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCG 136
           | |||   || ||   |||    ||   | |   ||     ||    '|   ||  | |||
Db      64 GTCAGTGGGTGGGGATGGCGCGGGGGTTGTTGTCGGTTCCGGTGTTTGTGGAGTCGGTCG 123

Qy     137 GTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTG 196
              |  ||  ||      |||||||  | | |||||      || |||||||  |||   |
Db     124 TGGAGTGCGATGCGGTCGTGTCGAGCGTCGTCGGCTTCAGCGTGCTGGGCGTCCTGGAGG 183

Qy     197 GCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTGG 256
           ||   |    ||  | |||| ||||||     ||     || | | |  || ||   |||
Db     184 GCCGCAGCGGCGCCCCGAGCCTGGACCGCGTCGACGTGGTCCAGCCGGTCCTGTTCGTGG 243

Qy     257 AAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCG 290
             | || |     | ||   |    || ||| |
Db     244 TCATGGTCAGCCTGGCCCGCCTGTGGCGCTGGTG 277


RESULT 4
US-09-738-626-1633/c
; Sequence 1633, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1633
;   LENGTH: 810
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1633

  Query Match             7.0%;  Score 33.8;  DB 9;  Length 810;
  Best Local Similarity   56.9%;  Pred. No. 0.6;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      69 CTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCA 128
           | ||||| || ||   ||  |||||  | ||||||| ||   ||  | ||| | |  | |
Db     181 CGTCAGCAGCAAGCTTGTCTGCGCGCTCGCCGGTGCCGTCTGGGCTGTTGTCGTCCACGA 122

Qy     129 GCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCG 177
              || ||  || | |   ||| | |   ||   || |||  || || |
Db     121 TGAGTACGTGAACGTCAGGGGTTGCGGTGCGCACGCGATCCACGATGAG 73


RESULT 5
US-09-738-626-2633
; Sequence 2633, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
```

```
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2633
;   LENGTH: 1164
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-2633

  Query Match             7.0%;  Score 33.6;  DB 9;  Length 1164;
  Best Local Similarity   61.4%;  Pred. No. 0.82;
  Matches   54;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      55 GGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGC 114
           | | || ||  ||  ||| ||| |  ||| |   |  ||  || | |||    | | | |
Db     469 GATGGCACCTTGACTTTCCGCCACGTGGACGGGCGTACGTACATCAGTGGAAATATTGAC 528

Qy     115 GGTGTTGCCGGCCAGCTGTTCGGTAAAG 142
           |||||  ||||| ||||||||   ||||
Db     529 GGTGTCACCGGCAAGCTGTTCCAAAAAG 556


RESULT 6
US-09-780-996-6
; Sequence 6, Application US/09780996
; Patent No. US20020061553A1
; GENERAL INFORMATION:
;  APPLICANT: Maury, Isabella
;  APPLICANT:  Mercken, Luc
;  APPLICANT:  Fournier, Alain
;  TITLE OF INVENTION: Partners of the PTB1 Domain of FE65, Preparation and Uses
;  FILE REFERENCE: ST00004-US
;  CURRENT APPLICATION NUMBER: US/09/780,996
;  CURRENT FILING DATE:  2001-02-09
;  PRIOR APPLICATION NUMBER: FR00/01628
;  PRIOR FILING DATE: 2000-02-10
;  PRIOR APPLICATION NUMBER: US 60/198,500
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-780-996-6

  Query Match             6.9%;  Score 33.2;  DB 10;  Length 1047;
  Best Local Similarity   51.3%;  Pred. No. 1;
  Matches   77;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     297 GGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACA 356
           |||  |  | |  ||||| |||| || | ||     ||| |||      |||||
Db     834 GGAGATGGCTGATGGCTACGCTGTAGACCGGGCCATTACCCACCTCAACAACAACTTCAT 893

Qy     357 GGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAA 416
           |    | |   ||| | |   | | |  |   |||||  |||||| ||  | |||||||

Db     894 GTTTGGGCAGAAGCTGAATGTCTGTGTCTCCAAGCAGCCAGCCATCATGCCTGGTCAGTC 953

Qy     417 ACAGGAAATCTACGGCACCGCGTGCCCTCA 446
           | | |   |  | | |    | |||  | |
Db     954 ATACGGGTTGGAAGACGGGTCTTGCAGTTA 983


RESULT 7
US-10-086-738A-1
; Sequence 1, Application US/10086738A
; Publication No. US20030004306A1
; GENERAL INFORMATION:
;  APPLICANT: MIYAWAKI, Atsushi
;  APPLICANT:  NAGAI, Takeharu
;  TITLE OF INVENTION: A Fluorescent Protein
;  FILE REFERENCE: P22042
;  CURRENT APPLICATION NUMBER: US/10/086,738A
;  CURRENT FILING DATE:  2002-07-12
;  PRIOR APPLICATION NUMBER: PCT/JP05/98922001
;  PRIOR FILING DATE: 2001-03-05
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 1278
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Construct from several species
US-10-086-738A-1

  Query Match             6.9%;  Score 33.2;  DB 9;  Length 1278;
  Best Local Similarity   55.1%;  Pred. No. 1.1;
  Matches   65;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     337 TACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAA 396
           | | || |  ||| |     | | |||    ||| | |  ||  ||| ||  || || |
Db     733 TTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC 792

Qy     397 GCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACCGCGTGCCCTCAGCCGGATG 454
           | ||  |||  || || || | |||   | |||| |||| |  ||  | | | || |
Db     793 GGCAACATCCTGGGGCACAAGCTGGAGTACAACGGTACCGGGGACCAACTGACAGAAG 850


RESULT 8
US-09-822-830A-420
; Sequence 420, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
;  APPLICANT: Genetics Institute, Inc.
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6402
;  CURRENT APPLICATION NUMBER: US/09/822,830A
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,604
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 631
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 420
;   LENGTH: 1792
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-830A-420

  Query Match             6.9%;  Score 33.2;  DB 10;  Length 1792;
  Best Local Similarity   51.3%;  Pred. No. 1.3;
```

```
  Matches   77;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     297 GGACACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACA 356
           ||| |  | | |  ||||| |||| || | ||     ||| |||      |||||
Db    1214 GGAGATGGCTGATGGCTACGCTGTAGACCGGGCCATTACCCACCTCAACAACAACTTCAT 1273

Qy     357 GGAACGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAA 416
           |    | |   ||| | |   | | |  |   |||||  |||||| ||  | ||||||
Db    1274 GTTTGGGCAGAAGCTGAATGTCTGTGTCTCCAAGCAGCCAGCCATCATGCCTGGTCAGTC 1333

Qy     417 ACAGGAAATCTACGGCACCGCGTGCCCTCA 446
           | | |   |  | | |    | |||  | |
Db    1334 ATACGGGTTGGAAGACGGGTCTTGCAGTTA 1363


RESULT 9
US-09-712-363-88/c
; Sequence 88, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 88
;   LENGTH: 900
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-88

  Query Match             6.8%;  Score 32.8;  DB 9;  Length 900;
  Best Local Similarity   51.4%;  Pred. No. 1.3;
  Matches   76;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy      50 TGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTG 109
           ||||  |   |||  |      |  |  |     ||  || | ||  | |  ||  | |||||
Db     616 TGGTCGTGGGCGGCGGTGCCGCCGCCGTTGTGGTGGTCGCCGGCGGCGGCGCCGTGGTTG 557

Qy     110 TGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCG 169
           | | |||||  | |||          ||  | ||   | |   ||||  ||  | ||  |||
Db     556 TCGCCGGTGGGGGCGGTGGTGGCGGCGTCACAGTGACCGTCTGTGTCTCCGTCGCTGTCG 497

Qy     170 GCGGTGCGGTTCTGGGCGGTCTGATTGG 197
           |     ||||  |||||||| |  |||
Db     496 GGATCTCGGTGGTGGGCGGTGGGGCTGG 469


RESULT 10
US-10-115-195-3/c
; Sequence 3, Application US/10115195
; Patent No. US20020168736A1
; GENERAL INFORMATION:
;  APPLICANT: Brunkow, Mary E.
;  APPLICANT:  Jeffery, Eric W.
;  APPLICANT:  Hjerrild, Kathryn A.
;  APPLICANT:  Ramsdell, Fred
;  TITLE OF INVENTION: IDENTIFICATION OF THE GENE CAUSING THE
;  TITLE OF INVENTION:  MOUSE SCURFY PHENOTYPE AND ITS HUMAN ORTHOLOG
;  FILE REFERENCE: 240083.501C1
;  CURRENT APPLICATION NUMBER: US/10/115,195
;  CURRENT FILING DATE:  2002-04-02
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1869
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-115-195-3

  Query Match             6.6%;  Score 32;  DB 9;  Length 1869;
  Best Local Similarity   51.4%;  Pred. No. 3.3;
  Matches   74;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       6 TGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCA 65
           || ||| |||||||||     ||  |      | |||  |         | ||| || | |||
Db    1761 TGTTTGGCTGCAGGGCTCGACTGGGGGGTGTGTCTGGGGCGGAGGTGGGGGCTGGGGCCA 1702

Qy      66 GAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGG 125
           | ||    ||| |   ||          || ||  ||    | || ||| || | || | |
Db    1701 GGACCGGGGCCCCTCTGAGCAGCCTTGGGGCAAAGGATATGATGGGGGAGGGGGTGGCTG 1642

Qy     126 CCAGCTGTTCGGTAAAGGCTCTGG 149
           |||||  |      |  ||| ||||
Db    1641 CCAGCGGGGGAACAGGGGCCCTGG 1618


RESULT 11
US-10-057-609B-1/c
; Sequence 1, Application US/10057609B
; Publication No. US20030028919A1
; GENERAL INFORMATION:
;  APPLICANT: KARNOSKY, DAVID F.
;  APPLICANT:  PODILA, GOPI
;  APPLICANT:  XIANG, BIXIA
;  TITLE OF INVENTION: TRANSGENIC TREES HAVING INCREASED RESISTANCE TO
;  TITLE OF INVENTION:  IMIDAZOLINONE HERBICIDES
;  FILE REFERENCE: 16313-0093
;  CURRENT APPLICATION NUMBER: US/10/057,609B
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: 60/264,216
;  PRIOR FILING DATE: 2001-01-25
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2364
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-10-057-609B-1

  Query Match             6.6%;  Score 32;  DB 9;  Length 2364;
  Best Local Similarity   51.4%;  Pred. No. 3.6;
  Matches   74;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     101 GTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGA 160
           | | |  | |||    || |   ||  |  ||   |   ||||   ||||   |||  ||
Db     413 GGGAGAATCTGGAGATTGGTAATGGTGATTTGGAGGAGGAAGGAGATGGTTTGGTGGAGA 354

Qy     161 TGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGG 220
           ||  ||||  |  |  ||| | |  | | |  ||| | |    || | ||||   | |
Db     353 AGGAGATCGAAGAAGATGTTGTTGTTGTTGTTGTTGCCGCCGCCATGGTTCAGGAGAAAG 294

Qy     221 ACCAGCAGGATAAAATCAAACTGA 244
```

```
          |  ||||  ||  |  |    ||||  ||
Db    293 ATGAGCAAGAGATAGAGAAACAGA 270


RESULT 12
US-09-822-846-108/c
; Sequence 108, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Bowman, Michael R.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6400
;  CURRENT APPLICATION NUMBER: US/09/822,846
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,605
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 629
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
;   LENGTH: 2863
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-846-108

  Query Match             6.6%;  Score 32;  DB 9;  Length 2863;
  Best Local Similarity   51.4%;  Pred. No. 4;
  Matches   74;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy      6 TGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCA 65
          || ||| |||||||||    ||  |     | |||  |         | ||| || | |||
Db   2245 TGTTTGGCTGCAGGGCTCGACTGGGGGGTGTGTCTGGGGCGGAGGTGGGGGCTGGGGCCA 2186

Qy     66 GAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGG 125
          | ||    ||| |   ||          || ||  ||    | || ||| || | || | |
Db   2185 GGACCGGGGCCCCTCTGAGCAGCCTTGGGGCAAAGGATATGATGGGGGAGGGGGTGGCTG 2126

Qy    126 CCAGCTGTTCGGTAAAGGCTCTGG 149
          |||||  |      |  ||| ||||
Db   2125 CCAGCGGGGGAACAGGGGCCCTGG 2102


RESULT 13
US-09-379-931-6
; Sequence 6, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
;  APPLICANT:  Smit, John
;  APPLICANT:  Bingle, Wade H.
;  APPLICANT:  No. US20020009792A1ellini, John F.
;  TITLE OF INVENTION:  EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
;  FILE REFERENCE:  08106/002003
;  CURRENT APPLICATION NUMBER: US/09/379,931
;  CURRENT FILING DATE:  1999-08-24
;  PRIOR APPLICATION NUMBER:  US 08/614,377
;  PRIOR FILING DATE:  1996-03-12
;  PRIOR APPLICATION NUMBER:  US 08/194,290
;  PRIOR FILING DATE:  1994-02-09
;  PRIOR APPLICATION NUMBER:  US 07/895,367
;  PRIOR FILING DATE:  1992-06-09
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 3300
;   TYPE: DNA
;   ORGANISM: Caulobacter crescentus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)...(3178)
US-09-379-931-6

  Query Match             6.6%;  Score 32;  DB 10;  Length 3300;
  Best Local Similarity   48.4%;  Pred. No. 4.3;
  Matches   89;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy     24 CTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGA 83
          ||| |||| ||    | | |    ||||  |||   |||| | |||  ||    |   |
Db    856 CTCGCTGACCACCGGCACCGACACCCTGACGGGCACCGCCAACAACGACACGTTCGTTGC 915

Qy     84 AGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGG 143
            | ||  |  |||  || ||||    ||  || ||  | |  || |     ||||  ||
Db    916 GGGTGAAGTCGCCGGCGCTGCGACCCTGACCGTTGGCGACACCCTGAGCGGCGGTGCTGG 975

Qy    144 CTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAA 203
          | | |      ||     ||      |  | |||||||  ||     | ||  |||   |
Db    976 CACCGACGTCCTGAACTGGGTGCAAGCTGCTGCGGTTACGGCTCTGCCGACCGGCGTGAC 1035

Qy    204 AATC 207
          |||
Db   1036 GATC 1039


RESULT 14
US-09-738-626-2169/c
; Sequence 2169, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2169
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-2169

  Query Match             6.6%;  Score 31.8;  DB 9;  Length 891;
  Best Local Similarity   59.3%;  Pred. No. 2.7;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;
```

```
Qy      90 CGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGG 149
           ||| | | | |  ||| | |||||||   |||||| | |  |    ||   || || |||
Db     421 CGCCGTCGCGGACGCGTTCGTGGGCGCGTTTGCCGTCGATTTTGATGGCGGAGACTTTGG 362

Qy     150 TCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
           | |  | | |   |   ||| ||||| ||||
Db     361 TGGGTTTTGCATGATGTCGCCGGTGAGGTT 331


RESULT 15
US-09-963-690-1/c
; Sequence 1, Application US/09963690
; Patent No. US20020115160A1
; GENERAL INFORMATION:
;  APPLICANT: FARWICK, Mike et al.
;  TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE truB GENE
;  FILE REFERENCE: 032301 WD 229
;  CURRENT APPLICATION NUMBER: US/09/963,690
;  CURRENT FILING DATE:  2002-02-01
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1350
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (233)..(1123)
;   OTHER INFORMATION:
US-09-963-690-1

  Query Match             6.6%;  Score 31.8;  DB 10;  Length 1350;
  Best Local Similarity   59.3%;  Pred. No. 3.2;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      90 CGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGG 149
           ||| | | | |  ||| | |||||||   |||||| | |  |    ||   || || |||
Db     653 CGCCGTCGCGGACGCGTTCGTGGGCGCGTTTGCCGTCGATTTTGATGGCGGAGACTTTGG 594

Qy     150 TCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180
           | |  | | |   |   ||| ||||| ||||
Db     593 TGGGTTTTGCATGATGTCGCCGGTGAGGTT 563


Search completed: February 22, 2003, 19:00:08
Job time : 56.9931 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:11:08 ; Search time 2141.73 Seconds
(without alignments)
10435.915 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		295	38.4	489	6	AX252413	AX252413 Sequence
2		295	38.4	4983	1	AF184152	AF184152 Piscirick
3		270.2	35.2	486	6	AR001082	AR001082 Sequence
4	c	270.2	35.2	486	6	AR001083	AR001083 Sequence
5		270.2	35.2	486	6	AR027448	AR027448 Sequence
6	c	270.2	35.2	486	6	AR027449	AR027449 Sequence
7		270.2	35.2	486	6	AR058257	AR058257 Sequence
8	c	270.2	35.2	486	6	AR058258	AR058258 Sequence
9		270.2	35.2	486	6	I18514	I18514 Sequence 1
10	c	270.2	35.2	486	6	I18515	I18515 Sequence 3
11		270.2	35.2	486	6	I66614	I66614 Sequence 1
12	c	270.2	35.2	486	6	I66615	I66615 Sequence 3
13		270.2	35.2	486	6	I88789	I88789 Sequence 1
14	c	270.2	35.2	486	6	I88790	I88790 Sequence 3
15		270.2	35.2	499	6	AR096204	AR096204 Sequence
16		270.2	35.2	1146	6	AR151733	AR151733 Sequence
17		270.2	35.2	1146	6	AX250579	AX250579 Sequence
18		270.2	35.2	6760	1	CLOCBP	M73817 Clostridium
19		76.4	9.9	6240	1	AB004845	AB004845 Clostridi
20		70.8	9.2	4919	1	CCU40345	U40345 Clostridium
21		70.2	9.1	1069	12	AF283517	AF283517 Synthetic
22		70.2	9.1	1482	6	AR129965	AR129965 Sequence
23		70.2	9.1	1482	6	AR205057	AR205057 Sequence
24		70.2	9.1	2400	1	CTCIPBA	X68233 C.thermocel
25		70.2	9.1	5573	1	CLOCIPA	L08665 Clostridium
26		66.8	8.7	10898	1	AE007606	AE007606 Clostridi
27		61.6	8.0	6555	1	AF155197	AF155197 Acetivibr
28		51	6.6	3048	1	CLOFEEGASE	L04735 Clostridium
29		51	6.6	10782	1	CTH275974	AJ275974 Clostridi
30		47.4	6.2	3360	1	CSCELZ	X55299 C.stercorar
31		46.6	6.1	537	1	RIR17KGCA	D16515 Rickettsia
32		46	6.0	3067	1	CTCHIAGEN	Z68924 C.thermocel
33	c	45.8	6.0	8246	6	AX345104	AX345104 Sequence
34		45	5.9	3120	1	CSCELYZ	Z69359 C.stercorar
35		45	5.9	237523	1	RPXX04	AJ235273 Rickettsi
36	c	44.6	5.8	147192	2	AC116925	AC116925 Dictyoste
37		44.4	5.8	1011	1	CTH420770	AJ420770 Clostridi
38	c	44.4	5.8	10595	6	AX251295	AX251295 Sequence
39		44	5.7	3498	4	AY044905	AY044905 Canis fam
40		44	5.7	3632	6	AX082874	AX082874 Sequence
41	c	44	5.7	48839	2	AC095545	AC095545 Rattus no
42	c	44	5.7	125020	9	AF429315	AF429315 Homo sapi
43		44	5.7	174256	2	AC107099	AC107099 Rattus no
44	c	43.8	5.7	186622	2	AC023433	AC023433 Homo sapi
45		43.8	5.7	197164	9	AC064865	AC064865 Homo sapi

ALIGNMENTS

RESULT 1
AX252413
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 5 from Patent WO0168865.
ACCESSION AX252413
VERSION AX252413.1 GI:15985721
KEYWORDS .
SOURCE Piscirickettsia salmonis.
ORGANISM Piscirickettsia salmonis
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
Piscirickettsia.
REFERENCE 1 (bases 1 to 489)
AUTHORS Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and
Burzio,L.
TITLE Fish vaccine against piscirickettsia salmonis

```
  JOURNAL   Patent: WO 0168865-A 5 20-SEP-2001;
            Aqua Health (Europe) Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Piscirickettsia salmonis"
                     /db_xref="taxon:1238"
BASE COUNT      139 a     79 c    144 g    127 t
ORIGIN

  Query Match           38.4%;  Score 295;  DB 6;  Length 489;
  Best Local Similarity   76.0%;  Pred. No. 1e-62;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349
Db      8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy    350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409
Db     68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy    410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469
Db    128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy    470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529
Db    188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy    530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589
Db    248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy    590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649
Db    308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy    650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709
Db    368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy    710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
Db    428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 2
AF184152
LOCUS       AF184152                4983 bp    DNA     linear   BCT 01-OCT-2000
DEFINITION  Piscirickettsia salmonis alanine racemase (alr) gene, partial cds;
            and BAX (bax), DNA repair enzyme RadA (radA), 17 kDa antigen
            (ospA), and transposase (tnpA) genes, complete cds.
ACCESSION   AF184152
VERSION     AF184152.1  GI:10441343
KEYWORDS    .
SOURCE      Piscirickettsia salmonis.
  ORGANISM  Piscirickettsia salmonis
            Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
            Piscirickettsia.
REFERENCE   1  (bases 1 to 4983)
  AUTHORS   Kuzyk,M.A., Burian,J., Thornton,J.C. and Kay,W.W.
  TITLE     Identification of a genus-common Rickettsial surface antigen in the
            salmonid pathogen Piscirickettsia salmonis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4983)
  AUTHORS   Kuzyk,M.A. and Kay,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1999) Biochemistry & Microbiology, University of
            Victoria, Petch Building, PO Box 3055, Victoria, BC V8W 3P6, Canada
FEATURES             Location/Qualifiers
     source          1. .4983
                     /organism="Piscirickettsia salmonis"
                     /strain="LF-89"
                     /db_xref="taxon:1238"
                     /clone="pB12"
     source          3560. .4849
                     /organism="Piscirickettsia salmonis"
                     /strain="LF-89"
                     /db_xref="taxon:1238"
                     /clone="pB12"
                     /insertion_seq="insertion sequence-like element"
     gene            <1. .501
                     /gene="alr"
     CDS             <1. .501
                     /gene="alr"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="alanine racemase"
                     /protein_id="AAG17001.1"
                     /db_xref="GI:10441347"
                     /translation="NSAAIFNFSYERYDWVRPGIMLYGISPFADKNGVDLELQPVMHV
                     VSRLISVKQLRQGESVGYGATWQCPEDMQVGILSLGYGDGYPRLAASGTPFLVRGQRC
                     ALIGRVSMDMIAIDLRRCPDAGVGEAVTVWGQDLPVEEIARHVGTIAYELVCNMPLRA
                     PYIWQE"
     gene            complement(526. .1257)
                     /gene="bax"
     CDS             complement(526. .1257)
                     /gene="bax"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="BAX"
                     /protein_id="AAG17002.1"
                     /db_xref="GI:10441348"
                     /translation="MLSLAVITTAYLAQQPLFSAPTSKTQINTAYSQDIQSIKQEFII
                     KITQSENTINQIILQKRQHILSLIQKKSLSKQEITQLQALAKQYKLNNLHFKHPQDFS
                     LLLKRVDIIPNALISAQAINESNWGRSRFAVEGNNFFGMRCHYPGCGIIPKARPANNH
                     WEVANYSSMTASVRAYIHTLNTHNAYQALRDLRAHMRANHQDVSAFKLAEGLTAYSIK
                     GTKYVELIQTIITRYLSKESTSQIP"
     gene            1373. .2743
                     /gene="radA"
                     /evidence=experimental
     CDS             1373. .2743
                     /gene="radA"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA repair enzyme RadA"
                     /protein_id="AAG16998.1"
                     /db_xref="GI:10441344"
                     /translation="MSKQKNQYVCSDCGGIAKKWLGQCPHCQQWNSLSEVKEVLPNRP
                     GRSQRFAGFAGIDAPKVKCLSEITPEQISRQPVGICEFDRVLGGGLVHGAVILLGGDP
                     GIGKSTLLLQTSVNCTQFGKVLYVTGEESLEQVTLRSKRLGLSQDVDLRLLAETQVER
                     ILKAAEIEQPKVLIVDSIQTIFTESLQSAPGGVAQVRESAAILTQFAKRTGTCLFLVG
                     HVTKEGALAGPRVLEHMVDTVLYFEGEQDSRFRLLRAVKNRFGAANELGIFAMTETGL
                     KTVSNPSAIFLSRYEDLQPGSVVMVAWEGTRPLLVEVQALVDESHSSNPRRIAVGLDQ
                     QRLAMLLAVLNRHGGIASYDQDVFINVVGGMKITETAADLALLLACVSSLRGKALSQE
                     LIVFGEIGLSGEIRPVQRGLERIKEAVKHGFTQAIVPLANCPKQKVDGIEIVGVKHLE
                     DALATL"
     gene            2834. .3322
                     /gene="ospA"
                     /evidence=experimental
     CDS             2834. .3322
                     /gene="ospA"
                     /note="similar to the rickettsial 17 kDa antigen"
                     /codon_start=1
                     /transl_table=11
                     /product="17 kDa antigen"
                     /protein_id="AAG17000.1"
                     /db_xref="GI:10441346"
                     /translation="MNRGCLQGSSLIIISVFLVGCAQNFSRQEVGAATGAVVGGVAGQ
                     LFGKGSGRVAMAIGGAVLGGLIGSKIGQSMDQQDKIKLNQSLEKVKAGQVTRWRNPDT
                     GNSYSVEPVRTYQRYNKQERRQQYCREFQQKAMIAGQKQEIYGTACRQPDGRWQVIST
                     EK"
     repeat_region   3560. .3846
     gene            complement(3842. .4561)
```

```
                  /gene="tnpA"
     CDS          complement(3842. .4561)
                  /gene="tnpA"
                  /codon_start=1
                  /transl_table=11
                  /product="transposase"
                  /protein_id="AAG16999.1"
                  /db_xref="GI:10441345"
                  /translation="MDLTLISLFCVIDDFCQELLPQWNAILLEDTNKKRNKPSQMSTS
                  EIMTIMIYFHKSNYRNFKMYYLHVIKGSMVKYFPNSVSYNRFVELMPSILLPLCFFIA
                  AQGKTATGIYFVDSTILRVCHEKRASQNRAFKGLAKKSKSTMGWYYGFKLHIIVNDMG
                  ELMAFKMSKATTDDRVVLPKMAENLTGKIIGDKGYISQKLFDQLYEKGLQLITKIRKN
                  MKNKLVLMIDKILLKKRGIIA"
     repeat_region 4562. .4849
BASE COUNT     1362 a    778 c   1265 g   1578 t
ORIGIN

  Query Match           38.4%;  Score 295;  DB 1;  Length 4983;
  Best Local Similarity  76.0%;  Pred. No. 1.3e-62;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy     290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349
           | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||
Db    2841 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 2900

Qy     350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409
           ||||||| || || || ||||| || || || || || || ||||| |||||||||||| |
Db    2901 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 2960

Qy     410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469
           |||||||||| ||||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db    2961 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 3020

Qy     470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529
           | ||  | ||||| |||||||||||||    ||||| |||||||||||||| || || |
Db    3021 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 3080

Qy     530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589
           |||||  | ||||||| || ||||| || || || || |||||||||||||| || || |
Db    3081 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 3140

Qy     590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649
           | || || ||   ||| || || || || || |||||||||||||| || || || || |
Db    3141 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 3200

Qy     650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709
           | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db    3201 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 3260

Qy     710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
           | |||||||| || ||||  || |||||||| || ||||| || |||   || ||||||
Db    3261 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 3319


RESULT 3
AR001082
LOCUS       AR001082                 486 bp    DNA     linear   PAT 04-DEC-1998
DEFINITION  Sequence 1 from patent US 5738984.
ACCESSION   AR001082
VERSION     AR001082.1  GI:3963149
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O.
  TITLE     Kits and methods of detection using cellulose binding domain fusion
            proteins
  JOURNAL   Patent: US 5738984-A 1 14-APR-1998;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      194 a     89 c     79 g    124 t
ORIGIN

  Query Match           35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity  98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 4
AR001083/c
LOCUS       AR001083                 486 bp    DNA     linear   PAT 04-DEC-1998
DEFINITION  Sequence 3 from patent US 5738984.
ACCESSION   AR001083
VERSION     AR001083.1  GI:3963150
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O.
  TITLE     Kits and methods of detection using cellulose binding domain fusion
            proteins
  JOURNAL   Patent: US 5738984-A 3 14-APR-1998;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      124 a     79 c     89 g    194 t
ORIGIN

  Query Match           35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity  98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
```

```
RESULT 5
AR027448
LOCUS       AR027448                 486 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5856201.
ACCESSION   AR027448
VERSION     AR027448.1  GI:5938268
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Methods of detection using a cellulose binding domain fusion
            product
  JOURNAL   Patent: US 5856201-A 1 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      194 a     89 c     79 g    124 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 6
AR027449/c
LOCUS       AR027449                 486 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 3 from patent US 5856201.
ACCESSION   AR027449
VERSION     AR027449.1  GI:5938269
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Methods of detection using a cellulose binding domain fusion
            product
  JOURNAL   Patent: US 5856201-A 3 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      124 a     79 c     89 g    194 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||| | ||
Db     231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 7
AR058257
LOCUS       AR058257                 486 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5837814.
ACCESSION   AR058257
VERSION     AR058257.1  GI:5983834
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M. and Doi,R.
  TITLE     Cellulose binding domain proteins
  JOURNAL   Patent: US 5837814-A 1 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      194 a     89 c     79 g    124 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 8
AR058258/c
LOCUS       AR058258                 486 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 3 from patent US 5837814.
ACCESSION   AR058258
VERSION     AR058258.1  GI:5983835
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
          Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M. and Doi,R.
  TITLE     Cellulose binding domain proteins
  JOURNAL   Patent: US 5837814-A 3 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      124 a     79 c     89 g    194 t
ORIGIN

  Query Match           35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 9
I18514
LOCUS       I18514                   486 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 1 from patent US 5496934.
ACCESSION   I18514
VERSION     I18514.1  GI:1598869
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Nucleic acids encoding a cellulose binding domain
  JOURNAL   Patent: US 5496934-A 1 05-MAR-1996;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      194 a     89 c     79 g    124 t
ORIGIN

  Query Match           35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 10
I18515/c
LOCUS       I18515                   486 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 3 from patent US 5496934.
ACCESSION   I18515
VERSION     I18515.1  GI:1598870
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Nucleic acids encoding a cellulose binding domain
  JOURNAL   Patent: US 5496934-A 3 05-MAR-1996;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      124 a     79 c     89 g    194 t
ORIGIN

  Query Match           35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 11
I66614
LOCUS       I66614                   486 bp    DNA     linear   PAT 29-DEC-1997
DEFINITION  Sequence 1 from patent US 5670623.
ACCESSION   I66614
VERSION     I66614.1  GI:2724592
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Methods of use of cellulose binding domain proteins
  JOURNAL   Patent: US 5670623-A 1 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      194 a     89 c     79 g    124 t
ORIGIN
```

```
  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 12
I66615/c
LOCUS       I66615                   486 bp    DNA     linear   PAT 29-DEC-1997
DEFINITION  Sequence 3 from patent US 5670623.
ACCESSION   I66615
VERSION     I66615.1  GI:2724593
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Methods of use of cellulose binding domain proteins
  JOURNAL   Patent: US 5670623-A 3 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      124 a     79 c     89 g    194 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 13
I88789
LOCUS       I88789                   486 bp    DNA     linear   PAT 10-AUG-1998
DEFINITION  Sequence 1 from patent US 5719044.
ACCESSION   I88789
VERSION     I88789.1  GI:3408729
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Cellulose binding domain fusion proteins
  JOURNAL   Patent: US 5719044-A 1 17-FEB-1998;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      194 a     89 c     79 g    124 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 14
I88790/c
LOCUS       I88790                   486 bp    DNA     linear   PAT 10-AUG-1998
DEFINITION  Sequence 3 from patent US 5719044.
ACCESSION   I88790
VERSION     I88790.1  GI:3408730
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.
  TITLE     Cellulose binding domain fusion proteins
  JOURNAL   Patent: US 5719044-A 3 17-FEB-1998;
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="unknown"
BASE COUNT      124 a     79 c     89 g    194 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352
```

```
Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 15
AR096204
LOCUS       AR096204                499 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 5 from patent US 6005092.
ACCESSION   AR096204
VERSION     AR096204.1  GI:10024795
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Shoseyov,O. and Shani,Z.
  TITLE     Arabidopsis thaliana endo-1,4-.beta.-glucanase gene and promoter
  JOURNAL   Patent: US 6005092-A 5 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="unknown"
BASE COUNT      197 a     93 c     82 g    127 t
ORIGIN

  Query Match             35.2%;  Score 270.2;  DB 6;  Length 499;
  Best Local Similarity   98.9%;  Pred. No. 1.5e-56;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     21 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 80

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 140

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 200

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 260

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295


Search completed: February 22, 2003, 16:18:12
Job time : 2186.73 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 13:08:35 ; Search time 214.439 Seconds
(without alignments)
8065.404 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	768	100.0	768	22	AAF86248	DNA sequence of c1
	2	768	100.0	768	24	ABK52403	E. coli codon opti
	3	483	62.9	483	22	AAF86247	DNA sequence of E.
	4	483	62.9	483	24	ABK52402	DNA encoding Esche
	5	295	38.4	486	22	AAF86246	OspA 17kD antigen
	6	295	38.4	486	24	ABK52401	DNA encoding outer
	7	295	38.4	489	22	AAH79040	Piscirickettsia sa
	8	271.8	35.4	573	22	AAD11043	Clostridium cellul
	9	271.8	35.4	1030	22	AAD11044	Clostridium cellul
	10	270.2	35.2	486	15	AAQ72917	Cellulose binding
	11	270.2	35.2	486	20	AAV74072	C. cellulovorans C
	12	270.2	35.2	499	20	AAX24930	Clostridium cellul
	13	270.2	35.2	507	22	AAD11042	Clostridium cellul
	14	270.2	35.2	984	22	AAD11046	Chimeric S peptide
	15	270.2	35.2	1146	20	AAZ27550	Gaussia luciferase
	16	270.2	35.2	1146	24	AAD22201	Gaussia species CB
	17	270.2	35.2	1288	22	AAD11045	Clostridium cellul
c	18	118	15.4	118	22	AAF86254	PCR primer #4 used
c	19	118	15.4	118	24	ABK52409	Escherichia coli c
	20	108.4	14.1	110	22	AAF86252	PCR primer #2 used
	21	108.4	14.1	110	24	ABK52407	Escherichia coli c
c	22	102	13.3	102	22	AAF86255	PCR primer #5 used
c	23	102	13.3	102	24	ABK52410	Escherichia coli c
	24	92.4	12.0	94	22	AAF86253	PCR primer #3 used
	25	92.4	12.0	94	24	ABK52408	Escherichia coli c
	26	87	11.3	111	22	AAF86251	PCR primer #1 used
	27	87	11.3	111	24	ABK52406	Escherichia coli c
c	28	72.4	9.4	110	22	AAF86256	PCR primer #6 used
c	29	72.4	9.4	110	24	ABK52411	Escherichia coli c
	30	70.2	9.1	1438	21	AAZ45336	DNA encoding a man
	31	70.2	9.1	1482	20	AAX90978	DNA encoding pecta
	32	70.2	9.1	1482	20	AAZ31562	Pectate lyase CBD
	33	70.2	9.1	5562	19	AAT86625	C. thermocellum Ci
c	34	45.8	6.0	8246	24	ABL32202	Human immune syste
c	35	44.4	5.8	10595	22	AAS46541	Tumour suppressor
	36	44	5.7	3632	24	ABA94342	Canine cyclooxygen
c	37	42	5.5	866	22	AAI94068	Human neuroblastom
c	38	41.8	5.4	12409	24	AAS63312	Chemically pretrea
	39	41.6	5.4	390	22	AAF66075	Novel human polynu
	40	41.6	5.4	1270	24	ABK12293	cDNA encoding huma
	41	41.6	5.4	1662	22	AAI59301	Human polynucleoti
	42	41.6	5.4	1684	24	ABK50361	cDNA encoding embr
	43	41.6	5.4	1900	22	AAI93857	Human stomach canc
	44	41.6	5.4	1900	22	AAH15558	Human cDNA sequenc
	45	41.6	5.4	2364	21	AAZ33323	Human secreted pro

ALIGNMENTS

RESULT 1
AAF86248
ID AAF86248 standard; DNA; 768 BP.
XX
AC AAF86248;
XX
DT 11-JUL-2001 (first entry)
XX
DE DNA sequence of c17e2 ospA construct with N-terminal fusion partner.
XX
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; ds.
XX
OS Piscirickettsia salmonis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..768
FT /*tag= a
FT /partial
FT /product= "C17E2 OspA with N-terminal fusion partner"
FT /note= "No stop codon is given"
FT misc_feature 1..285
FT /*tag= b
FT /note= "DNA encoding undefined N-terminal fusion partner"
FT misc_feature 286..768
FT /*tag= c
FT /note= "Optimised OspA construct c17E2"
XX
PN CA2281913-A1.
XX

```
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81128.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 4; Fig 5; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents P. salmonis OspA DNA termed C17E2 optimised for
CC   expression in Escherichia coli fused to DNA encoding an undefined
CC   N-terminal fusion partner. The protein encoded by this fusion construct
CC   is used in a vaccine to create an anti-OspA antibody response.
XX
SQ   Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

  Query Match             100.0%;  Score 768;  DB 22;  Length 768;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-214;
  Matches  768;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAAGCCCAACATCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTG 300

Qy     301 CAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGC 360

Qy     361 CGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTC 420

Qy     421 GGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATT 480

Qy     481 GGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTG 540

Qy     541 GAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTAC 600

Qy     601 TCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTAC 660

Qy     661 TGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 720

Qy     721 GCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768


RESULT 2
ABK52403
ID   ABK52403 standard; DNA; 768 BP.
XX
AC   ABK52403;
XX
DT   12-AUG-2002  (first entry)
XX
DE   E. coli codon optimised OspA, 17e2 with N-terminal fusion DNA.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds; 17e2.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..768
FT                   /*tag=  a
FT                   /product= "17e2"
FT                   /note= "Escherichia coli codon optimised OspA with
FT                         N-terminal fusion"
FT                   /partial
FT                   /note= "No stop codon given"
FT   sig_peptide     1..285
FT                   /*tag=  a
FT                   /note= "This sequence represents the N-terminal fusion"
FT   mat_peptide     286..768
FT                   /*tag=  c
FT                   /label= "mature_17e2"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97869.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
```

```
XX
PS   Claim 16; Fig 5; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettsial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Escherichia coli codon optimised outer surface
CC   lipoprotein OspA (17e2) with an N-terminal fusion used in the creation of
CC   the vaccine described in the invention.
XX
SQ   Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

  Query Match             100.0%;  Score 768;  DB 24;  Length 768;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-214;
  Matches  768;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAAGCCCAACATCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTG 300

Qy     301 CAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGC 360

Qy     361 CGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTC 420

Qy     421 GGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATT 480

Qy     481 GGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGAACCAGTCTCTG 540

Qy     541 GAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCGGTAACAGCTAC 600

Qy     601 TCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCCGTCAGCAGTAC 660

Qy     661 TGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAATCTACGGCACC 720

Qy     721 GCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768


RESULT 3
AAF86247
ID   AAF86247 standard; DNA; 483 BP.
XX
AC   AAF86247;
XX
DT   11-JUL-2001  (first entry)
XX
DE   DNA sequence of E. coli optimised ospA gene 17E2.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; 17E2; ds.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..483
FT                   /*tag=  a
FT                   /partial
FT                   /product= "OspA"
FT                   /note= "Genus specific 17kDa antigen, the sequence does
FT                         not include a stop codon"
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81127.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Example 3; Fig 4C; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents DNA which has been optimised for expression in
CC   Escherichia coli to encode the P. salmonis OspA protein. An OspA protein
CC   with an N-terminal fusion partner is used in a vaccine to create an
CC   anti-OspA antibody response.
XX
SQ   Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

  Query Match             62.9%;  Score 483;  DB 22;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-131;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     286 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 345
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60
```

```
Qy     346 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 405
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120

Qy     406 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 465
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180

Qy     466 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 525
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240

Qy     526 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 585
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300

Qy     586 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360

Qy     646 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420

Qy     706 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 765
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480

Qy     766 AAA 768
           |||
Db     481 AAA 483
```

```
RESULT 4
ABK52402
ID   ABK52402 standard; DNA; 483 BP.
XX
AC   ABK52402;
XX
DT   12-AUG-2002  (first entry)
XX
DE   DNA encoding Escherichia coli codon optimised OspA, 17e2.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds; 17e2.
XX
OS   Piscirickettsia salmonis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..483
FT                   /*tag=  a
FT                   /product= "17e2"
FT                   /note= "Escherichia coli codon optimised OspA"
FT                   /partial
FT                   /note= "No stop codon given"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97868.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 16; Fig 4; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Escherichia coli codon optimised outer surface
CC   lipoprotein OspA (17e2) used in the creation of the vaccine described in
CC   the invention.
XX
SQ   Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;
```

```
  Query Match             62.9%;  Score 483;  DB 24;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-131;
  Matches  483;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     286 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 345
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCGTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGC 60

Qy     346 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 405
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTT 120

Qy     406 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 465
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTT 180

Qy     466 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 525
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGGGCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAA 240

Qy     526 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 585
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGAACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGAC 300

Qy     586 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACCGGTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAA 360

Qy     646 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CGCCGTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAG 420

Qy     706 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 765
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAAATCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAA 480

Qy     766 AAA 768
           |||
Db     481 AAA 483
```

```
RESULT 5
AAF86246
ID   AAF86246 standard; DNA; 486 BP.
XX
AC   AAF86246;
XX
```

```
DT   11-JUL-2001  (first entry)
XX
DE   OspA 17kD antigen gene.
XX
KW   Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW   vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;
KW   SRS; ds.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag=  a
FT                   /partial
FT                   /product= "OspA"
FT                   /note= "Genus specific 17kDa antigen, the sequence does
FT                         not include a stop codon"
XX
PN   CA2281913-A1.
XX
PD   17-MAR-2001.
XX
PF   17-SEP-1999;   99CA-2281913.
XX
PR   17-SEP-1999;   99CA-2281913.
XX
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2001-316844/34.
DR   P-PSDB; AAB81126.
XX
PT   Method for protecting poikilothermic fish against salmonid rickettsial
PT   septicaemia and other rickettsial diseases comprises administering a
PT   vaccine containing the OspA protein of Piscirickettsia salmonis  -
XX
PS   Disclosure; Fig 2B; 35pp; English.
XX
CC   This invention relates to a method for the protection against infection
CC   of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC   salmonis. The method comprises administering an immunogenic amount of a
CC   P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC   OspA in the form of a vaccine. The method is used for protecting animals,
CC   particularly poikilothermic fish, against the bacterial pathogen
CC   P. salmonis. The method is also useful for protecting against salmonid
CC   rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC   sequence represents DNA encoding the P. salmonis OspA protein. An OspA
CC   protein with an N-terminal fusion partner is used in a vaccine to create
CC   an anti-OspA antibody response.
XX
SQ   Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

  Query Match             38.4%;  Score 295;  DB 22;  Length 486;
  Best Local Similarity   76.0%;  Pred. No. 4e-76;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349
          | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||
Db      8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy    350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409
          ||||||| || || || ||||| || || || || || || ||||| |||||||||||| |
Db     68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy    410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469
          |||||||||| ||||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db    128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy    470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529
          | ||  | ||||| ||||||||||||||    ||||| |||||||||||||| || || |

Db    188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy    530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589
          |||||  | |||||| || ||||| || || || || ||||||||||||| || || |
Db    248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy    590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649
          | || || ||   ||| || || || || || |||||||||||||| || || || || |
Db    308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy    650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709
          | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db    368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy    710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
          | |||||||| || ||||  || |||||||| || ||||| || |||   || ||||||
Db    428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 6
ABK52401
ID   ABK52401 standard; DNA; 486 BP.
XX
AC   ABK52401;
XX
DT   12-AUG-2002  (first entry)
XX
DE   DNA encoding outer surface lipoprotein OspA.
XX
KW   Outer surface lipoprotein; OspA; antibacterial; immunosupressive;
KW   vaccine; poikilothermic fish; fin-fish; Rickettsial septicaemia;
KW   Rickettisial disease; gene; ds.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag=  a
FT                   /product= "OspA"
FT                   /note= "Outer surface lipoprotein"
FT                   /partial= "No stop codon given"
XX
PN   CA2339327-A1.
XX
PD   15-MAR-2002.
XX
PF   19-MAR-2001; 2001CA-2339327.
XX
PR   15-SEP-2000; 2000US-0677374.
XX
PA   (THOR/) THORNTON J C.
PA   (KAYW/) KAY W W.
PA   (BURI/) BURIAN J.
PA   (KUZY/) KUZYK M A.
XX
PI   Thornton JC,  Kay WW,  Burian J,  Kuzyk MA;
XX
DR   WPI; 2002-455221/49.
DR   P-PSDB; AAU97867.
XX
PT   Inducing immunity in fin fish to Rickettsial septicaemia, comprises
PT   administration of an outer surface lipoprotein (OspA) of a bacterial
PT   strain, as a vaccine -
XX
PS   Claim 19; Fig 2; 55pp; English.
XX
CC   The invention describes a method of protecting a poikilothermic fish
CC   against infection by the bacterial pathogen Piscirickettsia Salmonis
CC   comprising administering either intraperitoneally, by immersion or
CC   orally, an immunogenic amount of principal antigen, the OspA (outer
CC   surface lipoprotein), its variants, non-lipidated form or antigenic
CC   peptides derived or synthesized with or without an adjuvant. The new
CC   method is used to provide an outer surface lipoprotein (OspA) of
```

```
CC   bacterial strain Piscirickettsia salmonis as a vaccine to induce immunity
CC   in fin-fish against Rickettsial septicaemia and other related
CC   Rickettisial diseases caused by either a virus, bacteria or parasite.
CC   This sequence encodes the Piscirickettsia salmonis outer surface
CC   lipoprotein, OspA, used in the creation of the vaccine described in the
CC   invention.
XX
SQ   Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

  Query Match             38.4%;  Score 295;  DB 24;  Length 486;
  Best Local Similarity   76.0%;  Pred. No. 4e-76;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy     290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349
           | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||
Db       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy     350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409
           ||||||| || || || ||||| || || || || || || || ||||| ||||||||||| |
Db      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy     410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469
           |||||||||| |||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy     470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529
           | ||  | ||||| ||||||||||||||    ||||| |||||||||||||| || || |
Db     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy     530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589
           |||||  | ||||||| || ||||| || || || || || |||||||||||||| || || |
Db     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy     590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649
           | || || ||   ||| || || || || || ||||||||||||||| || || || || |
Db     308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy     650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709
           | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db     368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy     710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
           | |||||||| || |||| || |||||||| || ||||| || |||   || ||||||
Db     428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 7
AAH79040
ID   AAH79040 standard; cDNA; 489 BP.
XX
AC   AAH79040;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Piscirickettsia salmonis polynucleotide P10.6.
XX
KW   Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW   septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW   ATCC VR-1361; ss.
XX
OS   Piscirickettsia salmonis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..489
FT                   /*tag=  a
XX
PN   WO200168865-A2.
XX
PD   20-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-GB01055.
XX
PR   11-MAR-2000; 2000GB-0005838.
PR   01-JUL-2000; 2000GB-0016080.
PR   01-JUL-2000; 2000GB-0016082.
PR   29-JUL-2000; 2000GB-0018599.
XX
PA   (AQUA-) AQUA HEALTH EURO LTD.
XX
PI   Simard N,  Brouwers H,  Jones S,  Griffiths S,  Valenzuela P;
PI   Burzio L;
XX
DR   WPI; 2001-639050/73.
DR   P-PSDB; AAG78025.
XX
PT   New nucleic acids encoding an amino acid sequence homologous to the
PT   surface antigen present on Piscirickettsia salmonis are useful to
PT   protect fish against piscirickettsiosis -
XX
PS   Claim 3; Fig 4; 25pp; English.
XX
CC   The invention relates to nucleic acid sequences and the encoded protein
CC   of a least part of the surface antigen present on Piscirickettsia
CC   salmonis for production of a vaccine with antibacterial activity to
CC   protect fish against P. salmonis which causes piscirickettsiosis, also
CC   known as salmonid rickettsial septicaemia.
XX
SQ   Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other;

  Query Match             38.4%;  Score 295;  DB 22;  Length 489;
  Best Local Similarity   76.0%;  Pred. No. 4e-76;
  Matches  364;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy     290 GTGGTTGCCTGCAGGGCAGCTCTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCC 349
           | || ||  |||| || ||   ||| || ||||||  ||| ||  | || || || ||||
Db       8 GAGGATGTTTGCAAGGTAGTAGTCTAATTATTATCAGTGTGTTTTTAGTTGGCTGTGCCC 67

Qy     350 AGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGTTGCCG 409
           ||||||| || || || ||||| || || || || || || ||||| ||||||||||| |
Db      68 AGAACTTTAGTCGTCAAGAAGTCGGAGCTGCGACTGGGGCTGTTGTTGGCGGTGTTGCTG 127

Qy     410 GCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGTTCTGG 469
           |||||||||| |||||||   |||||| ||  | |||||||| || ||||| ||| |||
Db     128 GCCAGCTGTTTGGTAAAGGTAGTGGTCGAGTTGCAATGGCCATTGGTGGTGCTGTTTTGG 187

Qy     470 GCGGTCTGATTGGCTCTAAAATCGGTCAGAGCATGGACCAGCAGGATAAAATCAAACTGA 529
           | ||  | ||||| ||||||||||||||    ||||| |||||||||||||| || || |
Db     188 GTGGATTAATTGGTTCTAAAATCGGTCAATCGATGGATCAGCAGGATAAAATAAAGCTAA 247

Qy     530 ACCAGTCTCTGGAAAAAGTGAAAGCCGGCCAGGTTACTCGTTGGCGTAATCCGGACACCG 589
           |||||  | ||||||| || ||||| || || || || |||||||||||||| || || |
Db     248 ACCAGAGTTTGGAAAAGGTAAAAGCAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

Qy     590 GTAACAGCTACTCTGTGGAACCGGTTCGCACCTACCAGCGTTACAACAAACAGGAACGCC 649
           | || || ||   ||| || || || || || ||||||||||||||| || || || || |
Db     308 GCAATAGTTATAGTGTTGAGCCAGTGCGTACTTACCAGCGTTACAATAAGCAAGAGCGTC 367

Qy     650 GTCAGCAGTACTGCCGCGAATTTCAGCAGAAAGCCATGATCGCAGGTCAGAAACAGGAAA 709
           | ||||| || || || |||||||||| || || ||||| ||||| ||||| || || |
Db     368 GCCAGCAATATTGTCGAGAATTTCAGCAAAAGGCGATGATTGCAGGGCAGAAGCAAGAGA 427

Qy     710 TCTACGGCACCGCGTGCCCTCAGCCGGATGGCCGCTGGCAGGTGATTAGCACCGAAAAA 768
           | |||||||| || |||| || |||||||| || ||||| || |||   || ||||||
Db     428 TTTACGGCACTGCATGCCGGCAACCGGATGGTCGTTGGCAAGTCATTTCAACAGAAAAA 486


RESULT 8
AAD11043
ID   AAD11043 standard; DNA; 573 BP.
XX
AC   AAD11043;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Clostridium cellulovorans cellulose binding domain-180 (CBD-180) DNA.
XX
```

```
KW   Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW   yarn; fiber; textile; biological crosslinker; mechanical property;
KW   Clostridium cellulovorans cellulose binding domain; wet strength;
KW   durability; elasticity; CBDclos; cellulose binding protein A; CBP A;
KW   CBD-180; ds.
XX
OS   Clostridium cellulovorans.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..572
FT                   /*tag= a
FT                   /product= "Cellulose binding domain-180 protein"
FT                   /note= "CDS does not include stop codon"
FT                   /partial
XX
PN   WO200134091-A2.
XX
PD   17-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-IL00708.
XX
PR   08-NOV-1999;   99US-0164140.
PR   18-NOV-1999;   99US-0166389.
XX
PA   (CBDT-) CBD TECHNOLOGIES LTD.
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Levy I,  Nussinovitch A,  Shoseyov O;
XX
DR   WPI; 2001-457121/49.
DR   P-PSDB; AAE05746.
XX
PT   Preparation of a polysaccharide containing material having at least one
PT   desired structural, chemical, physical, electrical and/or mechanical
PT   property -
XX
PS   Example 1.2; Fig 1e-1g; 121pp; English.
XX
CC   The present invention relates to methods and compositions for cross-
CC   linking and/or modifying the properties of polysaccharide materials.
CC   The method involves treating the polysaccharide structure with a
CC   polysaccharide binding domain (PBD) fusion protein. The method is
CC   used to alter the structural, chemical, physical, electrical and
CC   mechanical properties of polysaccharide materials such as paper,
CC   yarns, fibers and textiles, using biological crosslinking agents.
CC   The polysaccharide containing materials have improved mechanical
CC   properties such as wet strengths, durability and elasticity. The PBD
CC   reagent is applied in the forming stage in fluting paper manufacture
CC   which eliminates the sizing step. The use of a biological crosslinker
CC   improves the recyclability of paper products. The PBD reagent maintains
CC   the fine fibers in a slurry therefore resulting in better recovery of
CC   raw materials. The PBD molecules are eluted by strong alkaline conditions
CC   which enhances the ability of the alkaline glue used in binding to
CC   penetrate paper.
CC   The present sequence is a DNA encoding Clostridium cellulovorans
CC   cellulose binding domain (CBD)-180. CBD-180 contains 180 amino acids of
CC   Clostridium cellulovorans CBD (CBDclos) of cellulose binding protein A
CC   (CBP A).
XX
SQ   Sequence 573 BP; 215 A; 112 C; 98 G; 148 T; 0 other;

  Query Match             35.4%;  Score 271.8;  DB 22;  Length 573;
  Best Local Similarity   99.3%;  Pred. No. 2.8e-69;
  Matches  273;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 62

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 122

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 182

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 242

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     243 GCAAGCCCAACATCAACCTATGATACATATGTTGA 277


RESULT 9
AAD11044
ID   AAD11044 standard; DNA; 1030 BP.
XX
AC   AAD11044;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Clostridium cellulovorans CBD cross linker protein (CCP) encoding DNA.
XX
KW   Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW   yarn; fiber; textile; biological crosslinker; mechanical property;
KW   wet strength; durability; elasticity; cellulose binding domain; CBD;
KW   CBD cross linker protein; CCP; ds.
XX
OS   Clostridium cellulovorans.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1028
FT                   /*tag= a
FT                   /product= "CCP protein"
FT                   /transl_except= "(pos: 1020..1028, aa:Pro-Asp)"
FT                   /note= "CDS does not include stop codon"
FT                   /partial
XX
PN   WO200134091-A2.
XX
PD   17-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-IL00708.
XX
PR   08-NOV-1999;   99US-0164140.
PR   18-NOV-1999;   99US-0166389.
XX
PA   (CBDT-) CBD TECHNOLOGIES LTD.
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Levy I,  Nussinovitch A,  Shoseyov O;
XX
DR   WPI; 2001-457121/49.
DR   P-PSDB; AAE05747.
XX
PT   Preparation of a polysaccharide containing material having at least one
PT   desired structural, chemical, physical, electrical and/or mechanical
PT   property -
XX
PS   Example 1.2; Fig 2b-2e; 121pp; English.
XX
CC   The present invention relates to methods and compositions for cross-
CC   linking and/or modifying the properties of polysaccharide materials.
CC   The method involves treating the polysaccharide structure with a
CC   polysaccharide binding domain (PBD) fusion protein. The method is
CC   used to alter the structural, chemical, physical, electrical and
CC   mechanical properties of polysaccharide materials such as paper,
CC   yarns, fibers and textiles, using biological crosslinking agents.
CC   The polysaccharide containing materials have improved mechanical
CC   properties such as wet strengths, durability and elasticity. The PBD
CC   reagent is applied in the forming stage in fluting paper manufacture
CC   which eliminates the sizing step. The use of a biological crosslinker
CC   improves the recyclability of paper products. The PBD reagent maintains
```

```
CC   the fine fibers in a slurry therefore resulting in better recovery of
CC   raw materials. The PBD molecules are eluted by strong alkaline conditions
CC   which enhances the ability of the alkaline glue used in binding to
CC   penetrate paper.
CC   The present sequence is a DNA encoding Clostridium cellulovorans
CC   cellulose binding domain (CBD) cross linker protein (CCP).
XX
SQ   Sequence 1030 BP; 402 A; 195 C; 166 G; 267 T; 0 other;

  Query Match             35.4%;  Score 271.8;  DB 22;  Length 1030;
  Best Local Similarity   99.3%;  Pred. No. 3.8e-69;
  Matches  273;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps*   0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 62

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 122

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 182

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 242

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     243 GCAAGCCCAACATCAACCTATGATACATATGTTGA 277


RESULT 10
AAQ72917
ID   AAQ72917 standard; DNA; 486 BP.
XX
AC   AAQ72917;
XX
DT   05-JUN-1995  (first entry)
XX
DE   Cellulose binding domain.
XX
KW   Cellulose binding domain; CBD; ds.
XX
OS   Clostridium cellulovorans.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag=  a
XX
PN   WO9424158-A.
XX
PD   27-OCT-1994.
XX
PF   14-APR-1994;   94WO-US04132.
XX
PR   14-APR-1993;   93US-0048164.
XX
PA   (REGC ) UNIV CALIFORNIA.
PA   (YISS ) YISSUM RES & DEV CO.
XX
PI   Doi RH,  Goldstein MA,  Shoseyov O,  Shpiegl I;
XX
DR   WPI; 1994-341767/42.
DR   P-PSDB; AAR62634.
XX
PT   Isolated cellulose binding domain and fusion proteins - with
PT   applications. including drug delivery, affinity separations. and
PT   diagnostic techniques
XX
PS   Claim 10; Fig 1; 125pp; English.
XX
CC   AAQ72917/R63634 is a novel isolated cellulose binding domain. It pref.
CC   binds crystalline with a Kd ranging from 1.5 microM - 0.5 microM,
CC   pref. with a Kd of <1.2 microM, esp. <1.0 microM. The protein and
CC   the encoding DNA are claimed, as is a CBD fusion protein (FP)
CC   comprising the CBD and a second protein. The second protein is pref.
CC   Protein A, heparinase, a hormone or an enzyme capable of degrading
CC   an environmental pollutant, or an HSP, HSP antibody, cross-reactive
CC   HSP-related protein or an antigenic portion of this. The CBD and FP
CC   may be used in drug delivery, affinity sepns. and diagnostic
CC   techniques. CBD nucleic acid may be obtd. from a variety of cell
CC   sources that produce CBDs that bind with high affinity and in a
CC   reversible manner or that produce CBD encoding mRNA. The preferred
CC   source of CBD encoding nucleic acid is C. cellulovorans.
XX
SQ   Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

  Query Match             35.2%;  Score 270.2;  DB 15;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 7.4e-69;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 11
AAV74072
ID   AAV74072 standard; DNA; 486 BP.
XX
AC   AAV74072;
XX
DT   09-MAR-1999  (first entry)
XX
DE   C. cellulovorans CbpA CBD DNA.
XX
KW   Cellulose binding domain; CBD; CbpA; fusion protein; diagnosis; HSP;
KW   immunoassay; heat-shock protein; cross reactive protein; detection;
KW   antigenic fragment; antibody; insulin-dependent diabetes mellitus;
KW   cellulose; chitin; cellulolytic; amorphogenic; ds.
XX
OS   Clostridium cellulovorans.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..486
FT                   /*tag=  a
FT                   /product= "cellulose binding domain"
XX
PN   US5856201-A.
XX
PD   05-JAN-1999.
XX
PF   27-OCT-1994;   94US-0330394.
XX
PR   27-OCT-1994;   94US-0330394.
PR   14-APR-1993;   93US-0048164.
PR   14-APR-1994;   94WO-US04132.
```

```
XX
PA   (REGC ) UNIV CALIFORNIA.
PA   (YISS ) YISSUM RES & DEV CO.
XX
PI   Doi RH,  Goldstein MA,  Shoseyov O,  Shpiegl I;
XX
DR   WPI; 1999-105130/09.
DR   P-PSDB; AAW90077.
XX
PT   Detection of a specific analyte by reaction with binding agent fused
PT   to cellulose binding domain - and subsequent treatment with
PT   cellulose and reaction of insoluble product with a label specific
PT   for the analyte
XX
PS   Example 7.2.1; Fig 1A-B; 63pp; English.
XX
CC   This sequence encodes a cellulose binding domain (CBD) derived from the
CC   CbpA protein of Clostridium cellulovorans. The sequence is used the
CC   construction of a fusion protein which can be used in diagnostic
CC   immunoassays, e.g. to detect heat-shock proteins (HSP) and their
CC   cross-reactive proteins, antigenic fragments or HSP-specific antibodies
CC   (which indicate insulin-dependent diabetes mellitus, or susceptibility
CC   to it. The CBD, has very high affinity for cellulose (including
CC   crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but
CC   has almost no cellulolytic or amorphogenic activities. The CBD binds over
CC   a wide pH range and is not released from cellulose by washing with water.
XX
SQ   Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

  Query Match            35.2%;  Score 270.2;  DB 20;  Length 486;
  Best Local Similarity  98.9%;  Pred. No. 7.4e-69;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 12
AAX24930
ID   AAX24930 standard; DNA; 499 BP.
XX
AC   AAX24930;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Clostridium cellulovorans cbpA cellulose binding domain DNA.
XX
KW   Cellulose binding domain; CBD; cbpA; endo-1,4-beta-glucanase; Cel1;
KW   Arabidopsis thaliana; transgenic plant; crop improvement;
KW   morphology; cell wall; ds.
XX
OS   Clostridium celulovorans.
XX
PN   WO9907830-A1.
XX
PD   18-FEB-1999.
XX
PF   26-JUL-1998;   98WO-IL00345.
XX
PR   13-JAN-1998;   98US-0006636.
PR   27-JUL-1997;   97IL-0121404.
PR   13-JAN-1998;   98US-0006632.
XX
PA   (YISS ) YISSUM RES & DEV CO.
XX
PI   Shani Z,  Shoseyov O,  Shpiegl E;
XX
DR   WPI; 1999-180488/15.
XX
PT   Transgenic plants expressing cell-wall modulating protein  - have
PT   altered morphology, e.g. increased growth, modified fiber length or
PT   cellulose content
XX
PS   Disclosure; Page 134; 144pp; English.
XX
CC   This DNA fragment encodes the cellulose binding domain (CBD) of
CC   the cbpA protein of Clostridium cellulovorans.  It was obtained
CC   by PCR amplification (see also AAX24952-53), and was used in the
CC   construction of binary vector pCC1, in which cbd was joined to
CC   the promoter and signal region of the novel endo-1,4-beta-glucanase
CC   cell gene (see AAX24923) of Arabidopsis thaliana.  Expression of cbd
CC   in transgenic tobacco plants modulated their growth.  The cell
CC   promoter can provide expression of any protein in elongating
CC   tissue.  CBD is an example of a cell wall modulation transgene
CC   used to alter the structure or morphology of a plant.  Transgenic
CC   plants of the invention may have altered biomass, growth, yield,
CC   greater or less resistance to biodegradation, be more or less
CC   digestible by ruminants, have altered cellulose content, larger or
CC   smaller leaves, etc., when compared to non-transgenic plants of the
CC   same species.
XX
SQ   Sequence 499 BP; 197 A; 93 C; 82 G; 127 T; 0 other;

  Query Match            35.2%;  Score 270.2;  DB 20;  Length 499;
  Best Local Similarity  98.9%;  Pred. No. 7.5e-69;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      21 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 80

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 140

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 200

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 260

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295


RESULT 13
AAD11042
ID   AAD11042 standard; DNA; 507 BP.
XX
AC   AAD11042;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Clostridium cellulovorans cellulose binding domain (CBDclos) DNA.
XX
KW   Polysaccharide modification; polysaccharide binding domain; PBD; paper;
```

```
KW   yarn; fiber; textile; biological crosslinker; mechanical property;
KW   Clostridium cellulovorans cellulose binding domain; wet strength;
KW   durability; elasticity; CBDclos; cellulose binding protein A; CBP A; ds.
XX
OS   Clostridium cellulovorans.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..494
FT                   /*tag= a
FT                   /product= "CBDclos protein"
XX
PN   WO200134091-A2.
XX
PD   17-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-IL00708.
XX
PR   08-NOV-1999;   99US-0164140.
PR   18-NOV-1999;   99US-0166389.
XX
PA   (CBDT-) CBD TECHNOLOGIES LTD.
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Levy I,  Nussinovitch A,  Shoseyov O;
XX
DR   WPI; 2001-457121/49.
DR   P-PSDB; AAE05745.
XX
PT   Preparation of a polysaccharide containing material having at least one
PT   desired structural, chemical, physical, electrical and/or mechanical
PT   property -
XX
PS   Example 1.1; Page 111; 121pp; English.
XX
CC   The present invention relates to methods and compositions for cross-
CC   linking and/or modifying the properties of polysaccharide materials.
CC   The method involves treating the polysaccharide structure with a
CC   polysaccharide binding domain (PBD) fusion protein. The method is
CC   used to alter the structural, chemical, physical, electrical and
CC   mechanical properties of polysaccharide materials such as paper,
CC   yarns, fibers and textiles, using biological crosslinking agents.
CC   The polysaccharide containing materials have improved mechanical
CC   properties such as wet strengths, durability and elasticity. The PBD
CC   reagent is applied in the forming stage in fluting paper manufacture
CC   which eliminates the sizing step. The use of a biological crosslinker
CC   improves the recyclability of paper products. The PBD reagent maintains
CC   the fine fibers in a slurry therefore resulting in better recovery of
CC   raw materials. The PBD molecules are eluted by strong alkaline conditions
CC   which enhances the ability of the alkaline glue used in binding to
CC   penetrate paper.
CC   The present sequence is a DNA encoding Clostridium cellulovorans
CC   cellulose binding domain (CBDclos) of cellulose binding protein A
CC   (CBP A).
XX
SQ   Sequence 507 BP; 200 A; 94 C; 84 G; 129 T; 0 other;

  Query Match            35.2%;  Score 270.2;  DB 22;  Length 507;
  Best Local Similarity  98.9%;  Pred. No. 7.6e-69;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      21 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 80

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 140

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 200

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 260

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295


RESULT 14
AAD11046
ID   AAD11046 standard; DNA; 984 BP.
XX
AC   AAD11046;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Chimeric S peptide-cellulose binding domain-S protein encoding DNA.
XX
KW   Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW   yarn; fiber; textile; biological crosslinker; mechanical property;
KW   wet strength; durability; elasticity; cellulose binding domain; CBD;
KW   chimeric protein; S peptide-cellulose binding domain-S protein;
KW   Spep-CBD-Sprot; bovine; ds.
XX
OS   Chimeric - Clostridium cellulovorans.
OS   Chimeric - Bos sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..984
FT                   /*tag= a
FT                   /product= "Spep-CBD-Sprot protein"
FT                   /transl_except= "(pos: 979..981, aa:Xaa)"
FT                   /note= "Xaa corresponds to in-frame stop codon;
FT                   CDS does not include start and stop codon"
FT                   /partial
FT   misc_feature    68..624
FT                   /*tag= b
FT                   /note= "This region is derived from C. cellulovorans"
FT   misc_feature    652..981
FT                   /*tag= c
FT                   /note= "This region is derived from bovine"
XX
PN   WO200134091-A2.
XX
PD   17-MAY-2001.
XX
PF   02-NOV-2000; 2000WO-IL00708.
XX
PR   08-NOV-1999;   99US-0164140.
PR   18-NOV-1999;   99US-0166389.
XX
PA   (CBDT-) CBD TECHNOLOGIES LTD.
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Levy I,  Nussinovitch A,  Shoseyov O;
XX
DR   WPI; 2001-457121/49.
DR   P-PSDB; AAE05749.
XX
PT   Preparation of a polysaccharide containing material having at least one
PT   desired structural, chemical, physical, electrical and/or mechanical
PT   property -
XX
PS   Example 4; Fig 4b-4g; 121pp; English.
XX
CC   The present invention relates to methods and compositions for cross-
CC   linking and/or modifying the properties of polysaccharide materials.
CC   The method involves treating the polysaccharide structure with a
CC   polysaccharide binding domain (PBD) fusion protein. The method is
CC   used to alter the structural, chemical, physical, electrical and
CC   mechanical properties of polysaccharide materials such as paper,
CC   yarns, fibers and textiles, using biological crosslinking agents.
CC   The polysaccharide containing materials have improved mechanical
```

```
CC   properties such as wet strengths, durability and elasticity. The PBD
CC   reagent is applied in the forming stage in fluting paper manufacture
CC   which eliminates the sizing step. The use of a biological crosslinker
CC   improves the recyclability of paper products. The PBD reagent maintains
CC   the fine fibers in a slurry therefore resulting in better recovery of
CC   raw materials. The PBD molecules are eluted by strong alkaline conditions
CC   which enhances the ability of the alkaline glue used in binding to
CC   penetrate paper.
CC   The present sequence is a DNA encoding S peptide-cellulose binding
CC   domain-S protein (Spep-CBD-Sprot), a fusion protein derived from
CC   Clostridium cellulovorans and bovine.
XX
SQ   Sequence 984 BP; 325 A; 240 C; 193 G; 226 T; 0 other;

  Query Match            35.2%;  Score 270.2;  DB 22;  Length 984;
  Best Local Similarity  98.9%;  Pred. No. 1.1e-68;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     103 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 162

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 222

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 282

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 342

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     343 GCAAGCCCAACATCAACCTATGATACATATGTTGA 377


RESULT 15
AAZ27550
ID   AAZ27550 standard; DNA; 1146 BP.
XX
AC   AAZ27550;
XX
DT   13-DEC-1999  (first entry)
XX
DE   Gaussia luciferase fusion protein coding sequence.
XX
KW   Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;
KW   bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;
KW   body paint; squirt gun; balloon; slimy play material; soap; toothpaste;
KW   fusion protein; ds.
XX
OS   Gaussia sp.
XX
PN   WO9949019-A2.
XX
PD   30-SEP-1999.
XX
PF   26-MAR-1999;   99WO-US06698.
XX
PR   27-MAR-1998;   98US-0079624.
PR   15-JUN-1998;   98US-0089367.
PR   01-OCT-1998;   98US-0102939.
XX
PA   (PROL-) PROLUME LTD.
PA   (BRYA/) BRYAN B J.
XX
PI   Bryan BJ,  Szent-Gyorgyi C;
XX
DR   WPI; 1999-580443/49.
DR   P-PSDB; AAY39952.
XX
PT   New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and
PT   Renilla and Ptilosarcus green fluorescent protein nucleic acids -
XX
PS   Disclosure; Page 222-223; 233pp; English.
XX
CC   This sequence encodes a luciferase of the invention. The invention
CC   relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and
CC   Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and
CC   proteins. The luciferases and GFPs can be used in
CC   bioluminescence-generating systems, assays, screening methods, diagnostic
CC   method and articles of manufacture. They can be expressed using
CC   e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla
CC   mulleri, Gaussia and  Pleuromamma luciferase or Renilla or Ptilosarcus
CC   GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,
CC   fairy dust, beverages, soft drinks, foods, textile products, bubbles,
CC   balloons, personal items, dentrifrices, soaps, body paints, bubble bath,
CC   ink or paper products. In particular, they can be used in e.g. squirt
CC   guns, pellet guns, finger paints, foot bags, greeting cards, slimy play
CC   material, clothing, bubble making toys, bath powders, cosmetics, body
CC   lotions, gels, body powders, body creams, toothpastes, mouthwashes,
CC   soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,
CC   frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,
CC   ice, dry ice or fountains. The nucleic acids can also be used to produce
CC   transgenic fish and plants.
XX
SQ   Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;

  Query Match            35.2%;  Score 270.2;  DB 20;  Length 1146;
  Best Local Similarity  98.9%;  Pred. No. 1.2e-68;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     241 GCAAGCCCAACATCAACCTATGATACATATGTTGA 275


Search completed: February 22, 2003, 14:45:04
Job time : 221.439 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:26:45 ; Search time 1523.18 Seconds
(without alignments)
8165.910 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	47	6.1	582	13	BJ429910	BJ429910 BJ429910
	2	43.4	5.7	551	9	AU037653	AU037653 AU037653
	3	43	5.6	1100	17	CNS00FCD	AL070512 Drosophil
	4	42.4	5.5	470	17	AZ799648	AZ799648 2M0057B14
c	5	42.2	5.5	976	17	CNS04E5M	AL286627 Tetraodon
	6	42	5.5	238	12	BG240789	BG240789 OV1_38_C0
	7	42	5.5	273	17	AQ260501	AQ260501 CITBI-E1-
	8	42	5.5	402	12	BG356895	BG356895 OV2_11_D0
	9	42	5.5	419	10	BE355894	BE355894 DG1_11_E1
	10	42	5.5	428	12	BF176742	BF176742 EM1_4_F01
	11	42	5.5	446	12	BG052290	BG052290 RHIZ2_12_
	12	42	5.5	500	12	BF586874	BF586874 FM1_31_E0
	13	42	5.5	507	12	BG053591	BG053591 RHIZ2_11_
	14	42	5.5	531	12	BG053014	BG053014 RHIZ2_16_
	15	42	5.5	537	12	BG713844	BG713844 EM1_20_C0
	16	42	5.5	561	12	BF587805	BF587805 FM1_40_G0
	17	42	5.5	577	12	BG673858	BG673858 EM1_20_C0
	18	42	5.5	582	10	BE356763	BE356763 DG1_12_B1
	19	42	5.5	591	10	BE356800	BE356800 DG1_12_E1
	20	42	5.5	597	12	BG102589	BG102589 RHIZ2_34_
	21	42	5.5	598	10	AW672446	AW672446 LG1_359_D
	22	42	5.5	600	10	BE360868	BE360868 DG1_67_A0
	23	42	5.5	620	12	BF481524	BF481524 FM1_21_B0
	24	42	5.5	649	10	BE355895	BE355895 DG1_11_E1
	25	41.8	5.4	238	9	AA352399	AA352399 EST60642
	26	41.8	5.4	474	13	BM031893	BM031893 497456 MA
	27	41.6	5.4	321	10	AW021656	AW021656 df26e07.y
	28	41.6	5.4	361	9	AA663164	AA663164 ac52h10.s
	29	41.6	5.4	459	13	BI493124	BI493124 df97h08.y
c	30	41.6	5.4	509	17	AZ447316	AZ447316 1M0244E04
	31	41.6	5.4	556	10	BE395015	BE395015 601309414
c	32	41.6	5.4	597	13	BI492596	BI492596 df26e07.w
	33	41.6	5.4	671	10	AW328753	AW328753 EST00024
	34	41.6	5.4	742	12	BF792098	BF792098 602252539
	35	41.6	5.4	878	13	BM449796	BM449796 AGENCOURT
c	36	41.6	5.4	958	17	CNS00IFF	AL074914 Drosophil
	37	41.6	5.4	1058	13	BM478067	BM478067 AGENCOURT
	38	41.6	5.4	1079	12	BF792265	BF792265 602252768
	39	41.6	5.4	1196	13	BM542460	BM542460 AGENCOURT
c	40	41.6	5.4	1201	17	CNS0167M	AL106396 Drosophil
	41	41.2	5.4	702	14	W27594	W27594 34h4 Human
	42	41.2	5.4	737	13	BM166352	BM166352 EST568875
	43	41	5.3	928	17	CNS00DKY	AL071865 Drosophil
	44	41	5.3	1249	14	BM909360	BM909360 AGENCOURT
c	45	40.6	5.3	942	17	CNS018GS	AL109318 Drosophil

ALIGNMENTS

RESULT 1
BJ429910/c
LOCUS BJ429910 582 bp mRNA linear EST 13-MAR-2002
DEFINITION BJ429910 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv5i09 3', mRNA sequence.
ACCESSION BJ429910
VERSION BJ429910.1 GI:19404632
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 582)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .582
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv5i09"

```
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      116 a     59 c     84 g    320 t      3 others
ORIGIN

  Query Match             6.1%;  Score 47;  DB 13;  Length 582;
  Best Local Similarity   47.7%;  Pred. No. 0.052;
  Matches  137;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy      10 GAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCAAAATT 69
           ||||   | |   |||   ||||| |   ||| |   ||| |||  |    |    ||
Db     435 GAATCAAATACAACTACAGAATCAACTACAACTACTACAACTACTACTTCTACAGCAACA 376

Qy      70 ACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATTACACA 129
           || |  ||| |||    | |   |||   |||  |    |||  || | || |  |  ||
Db     375 ACAACAACAACTGCAGATAAACCAAAGAAAAAGAAAAATAAAAATAAAAATAAAAAATCA 316

Qy     130 AGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTATTAGGAAAT 189
           | |     |||| ||  |   ||                 | ||  | ||  |  |  ||
Db     315 AATACAACTACAGAATCAACTACAACTACTACAACTACTACTTCTACAGCAACAACAACA 256

Qy     190 AGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCA 249
           |    ||  |||||  | |  |||  || |   ||        ||  ||| | |||  |||
Db     255 ACAACTGCAGATAAACCAAAGAAAAAGAAAAATAAAAATAAAAATAAAAATCAAAACCA 196

Qy     250 ACATCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTG 296
           | ||||||  || ||| | || | ||  |  |   ||    | ||||
Db     195 ATATCAACTGATAATAAAAATATAGAGTCAACAACTACTGATAGTTG 149


RESULT 2
AU037653
LOCUS       AU037653                551 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU037653 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSE138, mRNA sequence.
ACCESSION   AU037653
VERSION     AU037653.1  GI:3984406
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSE138"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      266 a     63 c     39 g    183 t
ORIGIN

  Query Match             5.7%;  Score 43.4;  DB 9;  Length 551;
  Best Local Similarity   52.5%;  Pred. No. 0.5;
  Matches   95;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy      32 CAGCACAAACAAACTCAATTACACCAATAATCAAAATTACTAACACATCTGACAGTGATT 91
           || |   ||||||    |||| |   ||||  ||||| | |||  ||    |     |||
Db      34 CAACTATAACAAATAATATTAAAATCATAAAAAAAAATCAGTAAATCAAAAAAATTAAATT 93

Qy      92 TAAATTTAAATGACGTAAAAGTTAGATATTATTACACAAGTGATGGTACACAAGGACAAA 151
           ||| | | ||    | |||   ||| ||| | |  ||| |  |  | ||  |   ||||
Db      94 CAAAATCACATATTTTTAAAAAAAGAAATTTTAAATCAAATTGTACTTCATCAAATCAAA 153

Qy     152 CTTTCTGGTGTGACCATGCTGGTGCATTATTAGGAAATAGCTATGTTGATAACACTAGCA 211
           ||       ||    |||   |  | | |||  ||| |   ||| | |||| | ||  |
Db     154 ATTCAAATAATGTAATTGCAACTTTAATTTTACCAAAAATTGATGATAATAATAATAATA 213

Qy     212 A 212
           |
Db     214 A 214


RESULT 3
CNS00FCD
LOCUS       CNS00FCD                1100 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR31D18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL070512
VERSION     AL070512.1  GI:4950453
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR31D18"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      356 a    100 c    111 g    137 t    396 others
ORIGIN

  Query Match             5.6%;  Score 43;  DB 17;  Length 1100;
  Best Local Similarity   33.7%;  Pred. No. 0.82;
  Matches   83;  Conservative    6;  Mismatches  157;  Indels    0;  Gaps    0;

Qy      25 AACAAATCAGCACAAACAAACTCAATTACACCAATAATCAAAATTACTAACACATCTGAC 84
           || ||   |  | ||| |||   ||  |   ||      | ||||   ||||
Db     104 AAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAANACATNNNNNATAATTTNNAACAATNTNANN 163

Qy      85 AGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATTACACAAGTGATGGTACACAA 144
           | ::   |||   ||  ||    |     | |:| :|  | | || : |    | | ||
```

```
Db    164 CGASMGGAAAAANAAAGGGAANNNANNNAAAAAHAAYAAMANATAAAWAAAAAAAAAAAA 223

Qy    145 GGACAAACTTTCTGGTGTGACCATGCTGGTGCATTATTAGGAAATAGCTATGTTGATAAC 204
          | | | ||           |       |       | | |   |       |
Db    224 CAAGACATTTANNNNNANTNNNNTNNNTNNNNANANNNNANANANANNNANNNANNAANN 283

Qy    205 ACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCAACATCAACCTATGAT 264
            |  |||   | |     |       |   | || ||     || |  ||      |
Db    284 NAAAAAAAANNNNANANNNNANNNNNNNNANNNANCAMCACNNNAAAAAAAAAANNNNNAA 343

Qy    265 ACATAT 270
          | |:||
Db    344 AAAKAT 349


RESULT 4
AZ799648
LOCUS       AZ799648               470 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0057B14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0057B14 F, DNA sequence.
ACCESSION   AZ799648
VERSION     AZ799648.1  GI:12950975
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0057  row: B  column: 14
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 470.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0057B14"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      138 a     97 c    132 g    103 t
ORIGIN

  Query Match             5.5%;  Score 42.4;  DB 17;  Length 470;
  Best Local Similarity   57.6%;  Pred. No. 0.89;
  Matches   76;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      2 TGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAA 61
          ||  || ||  |      |  ||||    | || ||||| |||||| ||  | ||||| ||
Db    297 TGGTAGCTGTTTGTGTTACATCTAGTTTACCACCACAACCAAACTAAACCAAACCAAAAA 356

Qy     62 TCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATT 121
          |||||| | | || |   |   ||   |  |||||    ||||  | || |||  || |
Db    357 TCAAAACTCCAAATAACCCAAGCATAAACATAAATGACCATGAGCTCAATGTTGTATTTG 416

Qy    122 ATTACACAAGTG 133
          ||  ||   |||
Db    417 ATGTCATTGGTG 428


RESULT 5
CNS04E5M/c
LOCUS       CNS04E5M               976 bp    DNA     linear   GSS 21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            103P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL286627
VERSION     AL286627.1  GI:8025084
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 976)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 976)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="103P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG103DH01LP1-end : T7"
BASE COUNT      243 a    151 c    137 g    327 t    118 others
ORIGIN

  Query Match             5.5%;  Score 42.2;  DB 17;  Length 976;
  Best Local Similarity   33.5%;  Pred. No. 1.3;
  Matches  116;  Conservative   61;  Mismatches  168;  Indels    1;  Gaps    1;
```

```
Qy      11 AATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCAAAATTA 70
           :||| || |: |   | | |: :    ||  ||| |  : :| :  :||||| ||| |||
Db     888 WATTATATAWTTTNTAAANAWAWWATAAATAAAAATWTWAWAWWWTWATAATAAAATTTA 829

Qy      71 CTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATTACACAA 130
             || | || :     | |||||| :|| ||| |   |||| | : |: | ::||    :
Db     828 TAAAAAAATTWATTTNTTATTTAATWTTTAATAATAAAAAA-TAWWAWTTAWWTATNATW 770

Qy     131 GTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTATTAGGAAATA 190
           : |:   | : :|  : ||| |:|      : :   |  |  |  :   |     ::| :
Db     769 AWAAWAWAATWTWAAAWAAAAATWTWAAWATWTWAATTAATATTATWAATTATATWWANW 710

Qy     191 GCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCAA 250
             || ||  |||   |     | : | |   || ::  ::|:| ||| :  |:    ||
Db     709 TAAATTTTATTAAWTWTNNATTAAWAATAAATAAAWWAHWWAWANAAAAWAWAWAAAAAA 650

Qy     251 CATCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTGCAGGGCAGCT 310
           :: ||   |:: |   |:: ::: :  :: : :   |:::: |: :: |  :|:  |
Db     649 WWWAAAAAAAKKKTTTTTHWNWKKKKKKKKKKRVGGGBKBBYTBTVBAGRGVGBASGGA 590

Qy     311 CTCTGATCATTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTT 356
           |  |       | ::|| : | ||||     |||  ||||
Db     589 GGGAGGACGGAGGGAKAGCKYTCTBAGGGGGTGATTTCCACCACTT 544


RESULT 6
BG240789
LOCUS       BG240789                238 bp    mRNA    linear   EST 15-FEB-2001
DEFINITION  OV1_38_C06.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG240789
VERSION     BG240789.1  GI:12775862
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 3
            High quality sequence stop: 230
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .238
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       63 a     54 c     70 g     51 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 238;
  Best Local Similarity   59.0%;  Pred. No. 0.91;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      79 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 138

Qy     405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     139 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 198

Qy     465 TC 466
           |
Db     199 CC 200


RESULT 7
AQ260501
LOCUS       AQ260501                273 bp    DNA     linear   GSS 24-OCT-1998
DEFINITION  CITBI-E1-2509D17.TR CITBI-E1 Homo sapiens genomic clone 2509D17,
            DNA sequence.
ACCESSION   AQ260501
VERSION     AQ260501.1  GI:3787025
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2509D17"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      113 a     43 c     47 g     70 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 17;  Length 273;
  Best Local Similarity   59.0%;  Pred. No. 0.95;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      23 CTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCAAAATTACTAACACATCTG 82
           |||| || ||   || | ||||||| |||||  ||||  ||||  | |||   |   |
Db     126 CTAAAAATTCCATTCAGAAAAACTCATTTACAATAATATAGAAAAGAATTAAGCAAGTGG 185

Qy      83 ACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATTACACAAGTGATGGTACAC 142
           | ||  | || |  || |  |    ||||| | ||||||||||      |  |   ||| |
Db     186 AAAGACAGTTCATCTTCACGGCTTGAAAAGCTTGATATTATTAGGATGGCAAGACTACGC 245

Qy     143 AA 144
           ||
```

```
Db    246 AA 247


RESULT 8
BG356895
LOCUS       BG356895                402 bp    mRNA    linear   EST 06-MAR-2001
DEFINITION  OV2_11_D01.g1_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG356895
VERSION     BG356895.1  GI:13238881
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 47
            High quality sequence stop: 391
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 2 (OV2)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      107 a     82 c    117 g     96 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 402;
  Best Local Similarity   59.0%;  Pred. No. 1.1;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db    140 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 199

Qy    405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
           | ||||||| |   | | ||| | || |   ||   ||||||||   | || |||||
Db    200 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 259

Qy    465 TC 466
           |
Db    260 CC 261


RESULT 9
BE355894
LOCUS       BE355894                419 bp    mRNA    linear   EST 20-JUL-2000
DEFINITION  DG1_11_E11.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355894
VERSION     BE355894.1  GI:9297451
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 16
            High quality sequence stop: 336
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      102 a     82 c    116 g    119 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 10;  Length 419;
  Best Local Similarity   59.0%;  Pred. No. 1.1;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     78 CGGCGAGAAGCTGACCGACTAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 137

Qy    405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
           | ||||||| |   | | ||| | || |   ||   ||||||||   | || |||||
Db    138 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 197

Qy    465 TC 466
           |
Db    198 CC 199


RESULT 10
BF176742
LOCUS       BF176742                428 bp    mRNA    linear   EST 31-OCT-2000
DEFINITION  EM1_4_F01.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BF176742
VERSION     BF176742.1  GI:11064652
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
```

```
          Email: mmpratt@uga.edu
          Sequences have been trimmed to exclude PolyA, vector and regions
          below Phred quality 16. The threshold for highest quality sequence
          is 20.
          Seq primer: PolyTMix
          High quality sequence start: 32
          High quality sequence stop: 427
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      103 a     80 c    119 g    126 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 428;
  Best Local Similarity   59.0%;  Pred. No. 1.1;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      58 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 117

Qy     405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     118 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 177

Qy     465 TC 466
           |
Db     178 CC 179


RESULT 11
BG052290
LOCUS       BG052290                446 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_12_G04.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG052290
VERSION     BG052290.1  GI:12506830
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 11
            High quality sequence stop: 440
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      101 a     81 c    124 g    140 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 446;
  Best Local Similarity   59.0%;  Pred. No. 1.1;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      51 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 110

Qy     405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
           | ||||||| |   |  | |||  | || |   ||   ||||||||   | || |||||
Db     111 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 170

Qy     465 TC 466
           |
Db     171 CC 172


RESULT 12
BF586874
LOCUS       BF586874                500 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  FM1_31_E06.g1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
            propinquum cDNA, mRNA sequence.
ACCESSION   BF586874
VERSION     BF586874.1  GI:11679198
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: floral-induced meristems
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 8
            High quality sequence stop: 495
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Floral-Induced Meristem 1 (FM1)"
                     /note="Organ: Floral-induced meristems; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; mature plants were placed in a growth chamber for
                     15 days with 16 hr darkness  and 8 hr light (flowering is
                     induced by short-day conditions); 16 days after being
                     returned  to the greenhouse under natural long days during
                     late April/early May, meristems were  harvested The
                     library was made from poly-A RNA in the cloning vector
                     lambda ZAP II. Clones to be sequenced were prepared by
                     mass excision."
BASE COUNT      111 a     94 c    141 g    154 t
ORIGIN
```

```
  Query Match             5.5%;  Score 42;  DB 12;  Length 500;
  Best Local Similarity   59.0%;  Pred. No. 1.2;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     89 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 148

Qy    405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
          | ||||||| |   | | ||| | || |   ||   |||||||   | || |||||
Db    149 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 208

Qy    465 TC 466
           |
Db    209 CC 210


RESULT 13
BG053591
LOCUS       BG053591                507 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_11_C03.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG053591
VERSION     BG053591.1  GI:12509440
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: T7
            High quality sequence start: 95
            High quality sequence stop: 507
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      109 a     92 c    142 g    164 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 507;
  Best Local Similarity   59.0%;  Pred. No. 1.2;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db     54 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 113

Qy    405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
          | ||||||| |   | | ||| | || |   ||   |||||||   | || |||||
Db    114 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 173

Qy    465 TC 466
           |
Db    174 CC 175


RESULT 14
BG053014
LOCUS       BG053014                531 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_16_D04.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG053014
VERSION     BG053014.1  GI:12508270
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 42
            High quality sequence stop: 527
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      122 a    101 c    152 g    154 t      2 others
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 531;
  Best Local Similarity   59.0%;  Pred. No. 1.2;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
          || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db    121 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 180

Qy    405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
          | ||||||| |   | | ||| | || |   ||   |||||||   | || |||||
Db    181 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 240

Qy    465 TC 466
           |
Db    241 CC 242


RESULT 15
BG713844
LOCUS       BG713844                537 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  EM1_20_C08.g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG713844
VERSION     BG713844.1  GI:14007794
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
```

```
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 62
            High quality sequence stop: 536
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      118 a     98 c    152 g    169 t
ORIGIN

  Query Match             5.5%;  Score 42;  DB 12;  Length 537;
  Best Local Similarity   59.0%;  Pred. No. 1.2;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     345 CGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTGT 404
           || | ||||  | | |  | |||| || | || |   | | ||| || ||  | || ||
Db      92 CGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGG 151

Qy     405 TGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGGT 464
            | ||||||| |   |  | |||  | || |   ||   |||||||   | || |||||
Db     152 CGACGGCCAGATCAACTATGAAGAGTTTGTTAAGGTTATGATGGCCAAGTGAGGAGCGGT 211

Qy     465 TC 466
           |
Db     212 CC 213


Search completed: February 22, 2003, 17:06:30
Job time : 1557.18 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 14:36:28 ; Search time 63.6684 Seconds
(without alignments)
3699.288 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		270.2	35.2	486	1	US-08-048-164A-1	Sequence 1, Appli
2	c	270.2	35.2	486	1	US-08-048-164A-3	Sequence 3, Appli
3		270.2	35.2	486	1	US-08-460-462-1	Sequence 1, Appli
4	c	270.2	35.2	486	1	US-08-460-462-3	Sequence 3, Appli
5		270.2	35.2	486	1	US-08-460-457-1	Sequence 1, Appli
6	c	270.2	35.2	486	1	US-08-460-457-3	Sequence 3, Appli
7		270.2	35.2	486	1	US-08-460-458-1	Sequence 1, Appli
8	c	270.2	35.2	486	1	US-08-460-458-3	Sequence 3, Appli
9		270.2	35.2	486	2	US-08-460-455-1	Sequence 1, Appli
10	c	270.2	35.2	486	2	US-08-460-455-3	Sequence 3, Appli
11		270.2	35.2	486	2	US-08-330-394A-1	Sequence 1, Appli
12	c	270.2	35.2	486	2	US-08-330-394A-3	Sequence 3, Appli
13		270.2	35.2	499	3	US-09-006-636-5	Sequence 5, Appli
14		270.2	35.2	499	4	US-09-006-632-5	Sequence 5, Appli
15		270.2	35.2	499	4	US-09-325-274-5	Sequence 5, Appli
16		270.2	35.2	1146	4	US-09-277-716-21	Sequence 21, Appl
17		270.2	35.2	1146	4	US-09-609-161B-21	Sequence 21, Appl
18		70.2	9.1	1482	4	US-09-198-956-9	Sequence 9, Appli
19		70.2	9.1	1482	4	US-09-198-955A-11	Sequence 11, Appl
20		70.2	9.1	1482	4	US-09-694-531-11	Sequence 11, Appl
21		70.2	9.1	1482	4	US-09-670-141-9	Sequence 9, Appli
22	c	40.8	5.3	168575	4	US-09-426-290-1	Sequence 1, Appli
23		40	5.2	2645	4	US-08-960-780-31	Sequence 31, Appl
24		40	5.2	2645	4	US-09-073-898-31	Sequence 31, Appl
25		40	5.2	2645	4	US-09-371-913A-1	Sequence 1, Appli
26		39.4	5.1	2004	1	US-08-471-033-6	Sequence 6, Appli
27		39.4	5.1	2004	2	US-08-471-044-6	Sequence 6, Appli
28		39.4	5.1	2004	2	US-08-463-483A-6	Sequence 6, Appli
29		39.4	5.1	2004	2	US-08-471-046A-6	Sequence 6, Appli
30		39.4	5.1	2004	2	US-08-470-566B-6	Sequence 6, Appli
31		39.4	5.1	2004	2	US-08-469-334-6	Sequence 6, Appli
32		39.4	5.1	2004	3	US-09-300-529-6	Sequence 6, Appli
33		39.4	5.1	2655	1	US-08-471-033-4	Sequence 4, Appli
34		39.4	5.1	2655	2	US-08-471-044-4	Sequence 4, Appli
35		39.4	5.1	2655	2	US-08-463-483A-4	Sequence 4, Appli
36		39.4	5.1	2655	2	US-08-471-046A-4	Sequence 4, Appli
37		39.4	5.1	2655	2	US-08-470-566B-4	Sequence 4, Appli
38		39.4	5.1	2655	2	US-08-469-334-4	Sequence 4, Appli
39		39.4	5.1	2655	3	US-09-300-529-4	Sequence 4, Appli
40		39.4	5.1	4041	1	US-08-471-033-22	Sequence 22, Appl
41		39.4	5.1	4041	2	US-08-471-044-22	Sequence 22, Appl
42		39.4	5.1	4041	2	US-08-463-483A-22	Sequence 22, Appl
43		39.4	5.1	4041	2	US-08-471-046A-22	Sequence 22, Appl
44		39.4	5.1	4041	2	US-08-470-566B-22	Sequence 22, Appl
45		39.4	5.1	4041	2	US-08-469-334-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-048-164A-1
; Sequence 1, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-048-164A-1

Query Match 35.2%; Score 270.2; DB 1; Length 486;
Best Local Similarity 98.9%; Pred. No. 4.6e-73;
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 2
US-08-048-164A-3/c
; Sequence 3, Application US/08048164A
; Patent No. 5496934
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/048,164A
;      FILING DATE:  14-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-048-164A-3

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 3
US-08-460-462-1
; Sequence 1, Application US/08460462
; Patent No. 5670623
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,462
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..486
US-08-460-462-1

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
```

```
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 4
US-08-460-462-3/c
; Sequence 3, Application US/08460462
; Patent No. 5670623
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,462
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-460-462-3

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||

Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 5
US-08-460-457-1
; Sequence 1, Application US/08460457
; Patent No. 5719044
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN FUSION PROTEINS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,457
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..486
US-08-460-457-1

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 6
US-08-460-457-3/c
; Sequence 3, Application US/08460457
; Patent No. 5719044
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN FUSION PROTEINS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,457
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-460-457-3

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 7
US-08-460-458-1
; Sequence 1, Application US/08460458
; Patent No. 5738984
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    TITLE OF INVENTION:  KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAI
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,458
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..486
US-08-460-458-1

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 8
US-08-460-458-3/c
; Sequence 3, Application US/08460458
; Patent No. 5738984
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    TITLE OF INVENTION:  KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,458
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-460-458-3

  Query Match             35.2%;  Score 270.2;  DB 1;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 9
US-08-460-455-1
; Sequence 1, Application US/08460455
; Patent No. 5837814
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN PROTEINS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,455
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..486
US-08-460-455-1

  Query Match             35.2%;  Score 270.2;  DB 2;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75
```

```
Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290


RESULT 10
US-08-460-455-3/c
; Sequence 3, Application US/08460455
; Patent No. 5837814
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN PROTEINS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,455
;      FILING DATE:  concurrently herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,164
;      FILING DATE:  14-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7809-009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-460-455-3

  Query Match             35.2%;  Score 270.2;  DB 2;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197


RESULT 11
US-08-330-394A-1
; Sequence 1, Application US/08330394A
; Patent No. 5856201
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Yosef, Karmey
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  METHODS OF DETECTION USING THE
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 AVENUE OF THE AMERICAS
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/330,394A
;      FILING DATE:  27-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CIP OF PCT/US94/04132
;      FILING DATE:  14-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MISROCK, S. LESLIE
;      REFERENCE/DOCKET NUMBER:  7809-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66441 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..486
US-08-330-394A-1

  Query Match             35.2%;  Score 270.2;  DB 2;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      16 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 75
```

```
Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 135

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 195

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 255

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290
```

```
RESULT 12
US-08-330-394A-3/c
; Sequence 3, Application US/08330394A
; Patent No. 5856201
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Yosef, Karmey
;    APPLICANT:  Shpiegl, Itai
;    APPLICANT:  Goldstein, Marc A.
;    APPLICANT:  Doi, Roy H.
;    TITLE OF INVENTION:  METHODS OF DETECTION USING THE
;    TITLE OF INVENTION:  CELLULOSE BINDING DOMAIN
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 AVENUE OF THE AMERICAS
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/330,394A
;      FILING DATE:  27-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CIP OF PCT/US94/04132
;      FILING DATE:  14-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MISROCK, S. LESLIE
;      REFERENCE/DOCKET NUMBER:  7809-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66441 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-330-394A-3

  Query Match             35.2%;  Score 270.2;  DB 2;  Length 486;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     471 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 412
```

```
Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 352

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 292

Qy     181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 232

Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           |||||||||||||||||||||||||||||| | ||
Db     231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197
```

```
RESULT 13
US-09-006-636-5
; Sequence 5, Application US/09006636
; Patent No. 6005092
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shani, Ziv
;    TITLE OF INVENTION:  ARABIDOPSIS THALIANA ENDO-1,4-BETA-
;    TITLE OF INVENTION:  GLUCANASE GENE, PROMOTER AND PROTEIN
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/006,636
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baldwin, Geraldine F.
;      REGISTRATION NUMBER:  31,232
;      REFERENCE/DOCKET NUMBER:  7809-019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-006-636-5

  Query Match             35.2%;  Score 270.2;  DB 3;  Length 499;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 80

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 140

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 200

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 260

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295


RESULT 14
US-09-006-632-5
; Sequence 5, Application US/09006632
; Patent No. 6184440
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shani, Ziv
;    APPLICANT:  Shpigel, Etai
;    TITLE OF INVENTION:  TRANSGENIC PLANTS OF ALTERED
;    TITLE OF INVENTION:  MORPHOLOGY
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/006,632
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baldwin, Geraldine F.
;      REGISTRATION NUMBER:  31,232
;      REFERENCE/DOCKET NUMBER:  7809-018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-006-632-5

  Query Match             35.2%;  Score 270.2;  DB 4;  Length 499;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 80

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 140

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 200

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 260

Qy    241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
          |||||||||||||||||||||||||||||| | ||
Db    261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295


RESULT 15
US-09-325-274-5
; Sequence 5, Application US/09325274
; Patent No. 6323023
;  GENERAL INFORMATION:
;    APPLICANT:  Shoseyov, Oded
;    APPLICANT:  Shani, Ziv
;    TITLE OF INVENTION:  ARABIDOPSIS THALIANA ENDO-1,4-BETA-
;    TITLE OF INVENTION:  GLUCANASE GENE, PROMOTER AND PROTEIN
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/325,274
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/006,636
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baldwin, Geraldine F.
;      REGISTRATION NUMBER:  31,232
;      REFERENCE/DOCKET NUMBER:  7809-019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-325-274-5

  Query Match             35.2%;  Score 270.2;  DB 4;  Length 499;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-73;
  Matches  272;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ATGTCAGTTGAATTTTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 80

Qy     61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 140

Qy    121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTA 200

Qy    181 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 TTAGGAAATAGCTATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACA 260
```

```
Qy     241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275
           ||||||||||||||||||||||||||||| | ||
Db     261 GCAAGCCCAACATCAACCTATGATACATATGTTGA 295
```

Search completed: February 22, 2003, 18:48:24
Job time : 69.6684 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 16:18:29 ; Search time 74.7219 Seconds
(without alignments)
5772.590 Million cell updates/sec

Title: US-09-677-374-5
Perfect score: 768
Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	70.2	9.1	1482	12	US-10-072-152-11	Sequence 11, Appl
	2	44	5.7	3509	10	US-09-919-060-4	Sequence 4, Appli
c	3	44	5.7	3509	10	US-09-919-060-6	Sequence 6, Appli
	4	41.6	5.4	3195	12	US-10-044-090-394	Sequence 394, App
	5	40	5.2	2645	10	US-09-850-351A-31	Sequence 31, Appl
	6	40	5.2	2645	10	US-09-967-805-1	Sequence 1, Appli
	7	38.8	5.1	289	9	US-09-854-133-533	Sequence 533, App
	8	38.8	5.1	289	10	US-09-738-973-533	Sequence 533, App
c	9	38.4	5.0	500	10	US-09-864-761-11900	Sequence 11900, A
c	10	37.4	4.9	480	10	US-09-960-352-5301	Sequence 5301, Ap
	11	36	4.7	7791	10	US-09-764-877-3700	Sequence 3700, Ap
c	12	35.6	4.6	2481	10	US-09-729-674-141	Sequence 141, App
	13	35.2	4.6	22484	10	US-09-875-114-2	Sequence 2, Appli
	14	35.2	4.6	22484	10	US-09-880-107-3341	Sequence 3341, Ap
	15	35.2	4.6	335913	9	US-09-754-853A-2	Sequence 2, Appli
	16	35.2	4.6	335913	9	US-09-754-853A-3	Sequence 3, Appli
	17	35	4.6	70768	9	US-10-135-322-13	Sequence 13, Appl
c	18	34.8	4.5	413	10	US-09-960-352-2919	Sequence 2919, Ap
	19	34.8	4.5	426	9	US-10-125-815-6	Sequence 6, Appli
c	20	34.8	4.5	1138	9	US-09-938-842A-4079	Sequence 4079, Ap
c	21	34.8	4.5	2002	10	US-09-887-576-298	Sequence 298, App
c	22	34.8	4.5	2004	10	US-09-887-576-299	Sequence 299, App
c	23	34.8	4.5	640681	10	US-09-790-988-1	Sequence 1, Appli
	24	34.6	4.5	4059	9	US-09-784-554B-1	Sequence 1, Appli
	25	34.4	4.5	1720	10	US-09-887-576-583	Sequence 583, App
c	26	34.4	4.5	3694	10	US-09-880-107-1616	Sequence 1616, Ap
	27	34.4	4.5	7311	12	US-10-044-090-455	Sequence 455, App
c	28	34.4	4.5	9172	10	US-09-070-927A-65	Sequence 65, Appl
	29	34.4	4.5	13737	9	US-10-074-279-10	Sequence 10, Appl
	30	34.2	4.5	1207	9	US-09-938-842A-5018	Sequence 5018, Ap
c	31	33.8	4.4	810	9	US-09-738-626-1633	Sequence 1633, Ap
c	32	33.8	4.4	45845	10	US-09-927-091-6	Sequence 6, Appli
c	33	33.8	4.4	180557	12	US-10-003-806-6	Sequence 6, Appli
c	34	33.8	4.4	180557	12	US-10-003-806-9	Sequence 9, Appli
	35	33.6	4.4	1164	9	US-09-738-626-2633	Sequence 2633, Ap
c	36	33.6	4.4	3111	10	US-09-815-242-4518	Sequence 4518, Ap
c	37	33.6	4.4	3198	10	US-09-815-242-8519	Sequence 8519, Ap
c	38	33.6	4.4	9767	10	US-09-070-927A-130	Sequence 130, App
	39	33.4	4.3	2000	9	US-09-938-842A-3270	Sequence 3270, Ap
	40	33.2	4.3	416	10	US-09-960-352-4584	Sequence 4584, Ap
	41	33.2	4.3	1047	10	US-09-780-996-6	Sequence 6, Appli
	42	33.2	4.3	1278	9	US-10-086-738A-1	Sequence 1, Appli
	43	33.2	4.3	1792	10	US-09-822-830A-420	Sequence 420, App
c	44	33.2	4.3	2187	9	US-09-870-759-81	Sequence 81, Appl
	45	33.2	4.3	9172	10	US-09-070-927A-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-10-072-152-11
; Sequence 11, Application US/10072152
; Patent No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438Alel Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Query Match 9.1%; Score 70.2; DB 12; Length 1482;
Best Local Similarity 63.2%; Pred. No. 2.3e-10;

```
  Matches  108;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy      2 TGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAA 61
          ||   |||||||||||||||    ||    ||||     || ||||||| |   ||
Db   1004 TGAAGGTTGAATTCTACAACAGCAATCCTTCAGATACTACTAACTCAATCAATCCTCAGT 1063

Qy     62 TCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATT 121
          ||||  ||||||| ||      |||||   |  ||||     |  | | | | |||||||
Db   1064 TCAAGGTTACTAATACCGGAAGCAGTGCAATTGATTTGTCCAAACTCACATTGAGATATT 1123

Qy    122 ATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTG 172
          ||||  |||   || ||     |||  || || ||||||||||||||||||
Db   1124 ATTATACAGTAGACGGACAGAAAGATCAGACCTTCTGGTGTGACCATGCTG 1174


RESULT 2
US-09-919-060-4
; Sequence 4, Application US/09919060
; Patent No. US20020064845A1
; GENERAL INFORMATION:
;  APPLICANT: Wisnewski, Nancy
;  APPLICANT:  Brandt, Kevin S.
;  TITLE OF INVENTION: CANINE COX-1 AND COX-2 NUCLEIC ACID MOLECULES, PROTEINS AND USES
;  FILE REFERENCE: AD-1
;  CURRENT APPLICATION NUMBER: US/09/919,060
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: 60/224,486
;  PRIOR FILING DATE: 2000-08-11
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 3509
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (53)..(1864)
;   OTHER INFORMATION:
US-09-919-060-4

  Query Match             5.7%;  Score 44;  DB 10;  Length 3509;
  Best Local Similarity   56.1%;  Pred. No. 0.014;
  Matches   83;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy     46 TCAATTACACCAATAATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGAC 105
          | ||||  ||      || |  |||| ||||  |   ||   || |  | |||||||||
Db   2080 TGAATTTCAGTTTATATTATTATTATGAACAAAAGCGAAGATGTTGGATATTTAAATGCT 2139

Qy    106 GTAAAAGTTAGATATTATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGAC 165
          ||   ||   || |  ||    |||||  |  |  |||    |    ||||| |||| |
Db   2140 GTTCCAGGATGACAAAATGCTGCAAGTTTTTTTCGACACTATCGGGATTTCTAGTGTATC 2199

Qy    166 CATGCTGGTGCATTATTAGGAAATAGCT 193
            ||||||||||  || || || ||
Db   2200 TTCCCTGGTGCATTAGAAGCAACTACCT 2227


RESULT 3
US-09-919-060-6/c
; Sequence 6, Application US/09919060
; Patent No. US20020064845A1
; GENERAL INFORMATION:
;  APPLICANT: Wisnewski, Nancy
;  APPLICANT:  Brandt, Kevin S.
;  TITLE OF INVENTION: CANINE COX-1 AND COX-2 NUCLEIC ACID MOLECULES, PROTEINS AND USES
;  FILE REFERENCE: AD-1
;  CURRENT APPLICATION NUMBER: US/09/919,060
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: 60/224,486
;  PRIOR FILING DATE: 2000-08-11
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 3509
;   TYPE: DNA
;   ORGANISM: Canis familiaris
US-09-919-060-6

  Query Match             5.7%;  Score 44;  DB 10;  Length 3509;
  Best Local Similarity   56.1%;  Pred. No. 0.014;
  Matches   83;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy     46 TCAATTACACCAATAATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGAC 105
          | ||||  ||      || |  |||| ||||  |   ||   || |  | |||||||||
Db   1430 TGAATTTCAGTTTATATTATTATTATGAACAAAAGCGAAGATGTTGGATATTTAAATGCT 1371

Qy    106 GTAAAAGTTAGATATTATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGAC 165
          ||   ||   || |  ||    |||||  |  |  |||    |    ||||| |||| |
Db   1370 GTTCCAGGATGACAAAATGCTGCAAGTTTTTTTCGACACTATCGGGATTTCTAGTGTATC 1311

Qy    166 CATGCTGGTGCATTATTAGGAAATAGCT 193
            ||||||||||  || || || ||
Db   1310 TTCCCTGGTGCATTAGAAGCAACTACCT 1283


RESULT 4
US-10-044-090-394
; Sequence 394, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 394
;   LENGTH: 3195
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 201572.10
US-10-044-090-394

  Query Match             5.4%;  Score 41.6;  DB 12;  Length 3195;
  Best Local Similarity   52.3%;  Pred. No. 0.068;
  Matches   92;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    133 GATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTATTAGGAAATAGC 192
          || | | |||  |||| | || ||  | |||| |  ||    | || || |  ||||
Db    577 GAAGATCCACTTGGACCTAATTGCTATTATGACAAAACTAAATCCTTCTTTGATAATATT 636

Qy    193 TATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCAACA 252
          | |  |||| ||| |||  ||  || | ||| ||  | | ||||||  |  |       |
Db    637 TCTTGTGATGACAATAGAGAACGGAGACCAACCTGGGCTGAAGAAAGAAGATTAAATGCT 696

Qy    253 TCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTGCAGGGCAG 308
            ||| | ||  |     ||   |||    | |   |||||  |       |||||
Db    697 GAAACATTTGGAATCCCACTTCGTCCAAACCGTGGCCGTGGGGGATACAGAGGCAG 752


RESULT 5
US-09-850-351A-31
; Sequence 31, Application US/09850351A
; Patent No. US20020100080A1
;   GENERAL INFORMATION:
;        APPLICANT: Feitelson, Jerald S.
;                   Schnepf, H. Ernest
;                   Narva, Kenneth E.
;                   Stockhoff, Brian A.
;                   Schmeits, James
;                   Loewer, David
;                   Dullum, Charles Joseph
;                   Muller-Cohn, Judy
```

```
;                  Stamp, Lisa
;                  Morrill, George
;       TITLE OF INVENTION: No. US20020100080A1el Pesticidal Toxins and Nucleotide
;                           Sequences Which Encode These Toxins
;       NUMBER OF SEQUENCES: 144
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
;            STREET: 2421 N.W. 41st Street, Suite A-1
;            CITY: Gainesville
;            STATE: FL
;            COUNTRY: US
;            ZIP: 32606-6669
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/850,351A
;            FILING DATE: 07-May-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 09/073,898
;            FILING DATE: 06-MAY-1998
;            APPLICATION NUMBER: US 08/960,780
;            FILING DATE: 30-OCT-1997
;            APPLICATION NUMBER: US 60/029,848
;            FILING DATE: 30-OCT-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Sanders, Jay M.
;            REGISTRATION NUMBER: 39,355
;            REFERENCE/DOCKET NUMBER: MA-708CD1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 352-375-8100
;            TELEFAX: 352-372-5800
;   INFORMATION FOR SEQ ID NO: 31:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2645 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       ORIGINAL SOURCE:
;            INDIVIDUAL ISOLATE: PS177C8
;       SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-850-351A-31

  Query Match             5.2%;  Score 40;  DB 10;  Length 2645;
  Best Local Similarity   60.4%;  Pred. No. 0.18;
  Matches   64;  Conservative    1;  Mismatches   41;  Indels    0;  Gaps    0;

Qy      64 AAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTAT 123
           ||  | | | |||||  ||| ||||||  || ||  || || |  | | ||| || | ||
Db    2460 AACGTGAGTGACACACTTGAAAGTGATAAAATTTACAAGGATGGGACAATTAAATTTGAT 2519

Qy     124 TACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATG 169
           |  ||||   || ||: | | | ||||   || |  | ||||  ||
Db    2520 TTTACAAAATATAGTRAAAATGAACAAGGATTATTTTATGACAGTG 2565


RESULT 6
US-09-967-805-1
; Sequence 1, Application US/09967805
; Patent No. US20020120114A1
; GENERAL INFORMATION:
;  APPLICANT: Schnepf, Ernest H
;  APPLICANT:  Narva, Kenneth E
;  APPLICANT:  Stockhoff, Brian A
;  APPLICANT:  Finstad Lee, Stacey
;  APPLICANT:  Walz, Mikki
;  APPLICANT:  Sturgis, Blake
;  TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus
;  TITLE OF INVENTION:  Strains
;  FILE REFERENCE: MA-719XC2D1
;  CURRENT APPLICATION NUMBER: US/09/967,805
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: 09/371,913
;  PRIOR FILING DATE: 1999-08-10
;  PRIOR APPLICATION NUMBER: 60/095,955
;  PRIOR FILING DATE: 1998-08-10
;  PRIOR APPLICATION NUMBER: 60/138,251
;  PRIOR FILING DATE: 1999-06-08
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2645
;   TYPE: DNA
;   ORGANISM: Bacillus laterosporus
US-09-967-805-1

  Query Match             5.2%;  Score 40;  DB 10;  Length 2645;
  Best Local Similarity   60.4%;  Pred. No. 0.18;
  Matches   64;  Conservative    1;  Mismatches   41;  Indels    0;  Gaps    0;

Qy      64 AAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTAT 123
           ||  | | | |||||  ||| ||||||  || ||  || || |  | | ||| || | ||
Db    2460 AACGTGAGTGACACACTTGAAAGTGATAAAATTTACAAGGATGGGACAATTAAATTTGAT 2519

Qy     124 TACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGTGACCATG 169
           |  ||||   || ||: | | | ||||   || |  | ||||  ||
Db    2520 TTTACAAAATATAGTRAAAATGAACAAGGATTATTTTATGACAGTG 2565


RESULT 7
US-09-854-133-533
; Sequence 533, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.475C10
;  CURRENT APPLICATION NUMBER: US/09/854,133
;  CURRENT FILING DATE:  2001-05-11
;  NUMBER OF SEQ ID NOS: 735
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 533
;   LENGTH: 289
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-854-133-533

  Query Match             5.1%;  Score 38.8;  DB 9;  Length 289;
  Best Local Similarity   52.5%;  Pred. No. 0.11;
  Matches   85;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
           || ||  ||  |  | |      |||  || | ||||   | ||  || ||     || |
Db      60 ATTTCCTTTCTAGACAAAGGACCTAGAAATTGAGCATGCAAAACATCCATCCATTCATTC 119

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
           ||  |||| | || |  || | || || ||||||  |  |   || | |||   ||||
Db     120 ATTCAAATAATTAGCCAATTTTACCGTCATTTAATTCCACCAGAAGCAAATACTAGAATA 179

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGT 162
           | |   |  |||   |||| | ||  |   || || |  | |
Db     180 TCTAGAAGTAGTTTGGGTAAAGAAACATTTACATTTTAATAT 221
```

```
RESULT 8
US-09-738-973-533
; Sequence 533, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Indirias, Carol Yoseph
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Elliot, Mark
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Kalos, Michael D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.475C9
;  CURRENT APPLICATION NUMBER: US/09/738,973
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS:  587
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 533
;   LENGTH: 289
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-738-973-533

  Query Match             5.1%;  Score 38.8;  DB 10;  Length 289;
  Best Local Similarity   52.5%;  Pred. No. 0.11;
  Matches   85;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy       1 ATGTCAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATA 60
          || ||  ||  |  | |     |||  || | ||||   | ||  || ||     || |
Db      60 ATTTCCTTTCTAGACAAAGGACCTAGAAATTGAGCATGCAAAACATCCATCCATTCATTC 119

Qy      61 ATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATAT 120
          ||  |||| | || |  || | || || |||||| |  |   || | |||   ||||
Db     120 ATTCAAATAATTAGCCAATTTTACCGTCATTTAATTCCACCAGAAGCAAATACTAGAATA 179

Qy     121 TATTACACAAGTGATGGTACACAAGGACAAACTTTCTGGTGT 162
           | |   |  |||   |||| | ||  |   || || |  | |
Db     180 TCTAGAAGTAGTTTGGGTAAAGAAACATTTACATTTTAATAT 221


RESULT 9
US-09-864-761-11900/c
; Sequence 11900, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11900
;   LENGTH: 500
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC010492.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
US-09-864-761-11900

  Query Match             5.0%;  Score 38.4;  DB 10;  Length 500;
  Best Local Similarity   51.1%;  Pred. No. 0.2;
  Matches   90;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy     133 GATGGTACACAAGGACAAACTTTCTGGTGTGACCATGCTGGTGCATTATTAGGAAATAGC 192
           || | | |||  |||| | || ||  | ||||||  ||    | |  || |  ||||
Db     336 GAAGATGCACTTGGACCTAATTGCTTTTATGACCAAACTAAATCCTCCTTTGATAATATT 277

Qy     193 TATGTTGATAACACTAGCAAAGTGACAGCAAACTTCGTTAAAGAAACAGCAAGCCCAACA 252
           | || |||| ||| |||  ||  ||   ||| ||  | | ||| || |  |       |
Db     276 TCTGGTGATGACAATAGAGAACGGAGGCCAACCTGGGCTGAAGGAAGAAGATTAAATGCT 217

Qy     253 TCAACCTATGATACATATCTGGATCCGTCTCATATGCGTGGTTGCCTGCAGGGCAG 308
             ||| | ||  |    ||  |||    | |   | |||  |       |||||
Db     216 GAAACATTTGGAATCCCACTTTGTCCAAACCGTGGCCATGGGGGATACAGAGGCAG 161


RESULT 10
US-09-960-352-5301/c
; Sequence 5301, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
```

```
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5301
;   LENGTH: 480
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 23-LIB3057-018-Q1-K1-F3
US-09-960-352-5301

  Query Match             4.9%;  Score 37.4;  DB 10;  Length 480;
  Best Local Similarity   55.9%;  Pred. No. 0.39;
  Matches   71;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      6 AGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCAA 65
          | |  |||   | || | ||| |||  |  | ||| |||    |||| |   | ||| ||
Db    449 AATAAAATATAAAAAATTTAATAAAAAAAAAAAAAAAAAAAAATATTATAAATAAAATAAA 390

Qy     66 AATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATTA 125
          ||| | ||| | ||     | | |   | |  |||| |  |||||  || | || | |
Db    389 AATAAATAATAAATAAATAAATAAAAAATAAATAAAAAAAATAAAAAATAAAAATAAAAA 330

Qy    126 CACAAGT 132
           |||| |
Db    329 TACAAAT 323


RESULT 11
US-09-764-877-3700
; Sequence 3700, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3700
;   LENGTH: 7791
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3700

  Query Match             4.7%;  Score 36;  DB 10;  Length 7791;
  Best Local Similarity   51.2%;  Pred. No. 4.7;
  Matches   84;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    320 TTATCTCTGTTTTCCTGGTGGGTTGCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGG 379
          ||  || |  ||  |    |||| |||| ||    | |  |||    || ||  | ||
Db   1025 TTCCCTTTCCTTGGCCTTGGGGTGCGGCCCGGACGGGCGGGGGCCGAACAGGTGCGGGGG 1084

Qy    380 CCACCGGTGCGGTTGTGGGCGGTGTTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTG 439
          ||  ||| ||||  |  || ||    || ||    |||| | |  | |||   ||  | |
Db   1085 CCCGCGGGGCGGGGGAAGGGGGCCGGGCTGGAGGCCTGTGCAGCGAGGGCGAGGGCGGAG 1144

Qy    440 TGTCGATGGCCATCGGCGGTGCGGTTCTGGGCGGTCTGATTGGC 483
           |  ||  |   |||||||   | |    |  || ||||  | |
Db   1145 GGAGGAGCGGGCTCGGCGGGAGGTTGGGAGCAGGGCTGACAGCC 1188


RESULT 12
US-09-729-674-141/c
; Sequence 141, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6055-64X
;  CURRENT APPLICATION NUMBER: US/09/729,674
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/539,330
;  PRIOR FILING DATE: 2000-03-30
;  NUMBER OF SEQ ID NOS: 283
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 141
;   LENGTH: 2481
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-729-674-141

  Query Match             4.6%;  Score 35.6;  DB 10;  Length 2481;
  Best Local Similarity   52.7%;  Pred. No. 3.2;
  Matches   77;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy    344 GCGCCCAGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGGCGGTG 403
          || ||  |||  | | || | ||| |    ||  |||  |       ||| | || || |
Db    402 GCACCTGGAAGGTGACCCACAAGGTATAGAGCTGGGCGGCTTTCCTCGTTATAGGTGGAG 343

Qy    404 TTGCCGGCCAGCTGTTCGGTAAAGGCTCTGGTCGTGTGTCGATGGCCATCGGCGGTGCGG 463
          |    | |||| ||| ||  |   |  |  |||  ||| ||||| ||| |   ||  |||
Db    342 TCTTGGCCCAGATGTCCGAGAGCCGAGCATGTCCGGTGACGATGTCCACCACAGGGCCGG 283

Qy    464 TTCTGGGCGGTCTGATTGGCTCTAAA 489
          |  ||    |||  | | ||  ||
Db    282 TCAGGGCGCAGCTGTATTGGTCACAA 257


RESULT 13
US-09-875-114-2
; Sequence 2, Application US/09875114
; Patent No. US20020002131A1
; GENERAL INFORMATION:
;  APPLICANT: No. US20020002131A1thwestern University
;  APPLICANT:  No. US20020002131A11 Bouck
;  APPLICANT:  David Dawson
;  APPLICANT:  Paul Gillis
;  TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis
;  FILE REFERENCE: 0290-23U2
;  CURRENT APPLICATION NUMBER: US/09/875,114
;  CURRENT FILING DATE:  2001-06-06
;  PRIOR APPLICATION NUMBER: US 09/122,079
;  PRIOR FILING DATE: 1998-07-23
;  PRIOR APPLICATION NUMBER: PCT/US98/15228
;  PRIOR FILING DATE: 1998-07-23
;  PRIOR APPLICATION NUMBER: US 08/899,304
;  PRIOR FILING DATE: 1997-07-23
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 22484
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: Unsure
;   LOCATION: 1...22484
;   OTHER INFORMATION: "n" means either a, c, t, or g
US-09-875-114-2

  Query Match             4.6%;  Score 35.2;  DB 10;  Length 22484;
```

```
  Best Local Similarity   55.8%;  Pred. No. 14;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy       5 CAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCA 64
           |||   ||   | || ||| |||||| ||  |||||  ||   ||  |    || ||
Db    4908 CAGAGCAAGGTTCCATCTCAAACAAAACAAAACAAAACAAACAAAAAAACAAAACAAAAG 4967

Qy      65 AAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATT 124
             | ||| ||||||| |   |  ||   || | |||||    ||  | ||  || || ||
Db    4968 CTAATACAAACACATATACAATAGACAAAACTGTAAATATTTTATTATTTTTATTTTTTT 5027


RESULT 14
US-09-880-107-3341
; Sequence 3341, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3341
;   LENGTH: 22484
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 U29953
;   NAME/KEY: unsure
;   LOCATION: (1)..(22484)
;   OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3341

  Query Match             4.6%;  Score 35.2;  DB 10;  Length 22484;
  Best Local Similarity   55.8%;  Pred. No. 14;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy       5 CAGTTGAATTCTACAACTCTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCA 64
           |||   ||   | || ||| |||||| ||  |||||  ||   ||  |    || ||
Db    4908 CAGAGCAAGGTTCCATCTCAAACAAAACAAAACAAAACAAACAAAAAAACAAAACAAAAG 4967

Qy      65 AAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATT 124
             | ||| ||||||| |   |  ||   || | |||||    ||  | ||  || || ||
Db    4968 CTAATACAAACACATATACAATAGACAAAACTGTAAATATTTTATTATTTTTATTTTTTT 5027


RESULT 15
US-09-754-853A-2
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
;  APPLICANT: Hauge, Brian M.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Parsons, Jeremy D.
;  APPLICANT:  Wang, Ming Li
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Soybean Cyst Nematode Resistance
;  FILE REFERENCE: 38-10(15810)B
;  CURRENT APPLICATION NUMBER: US/09/754,853A
;  CURRENT FILING DATE:  2001-01-05
;  PRIOR APPLICATION NUMBER: US 60/174,880
;  PRIOR FILING DATE: 2000-01-07
;  NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
;   LENGTH: 335913
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)
;   OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-2

  Query Match             4.6%;  Score 35.2;  DB 9;  Length 335913;
  Best Local Similarity   55.8%;  Pred. No. 64;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      22 TCTAACAAATCAGCACAAACAAACTCAATTACACCAATAATCAAAATTACTAACACATCT 81
           ||  |  || ||  |||||||||||   | || |    || |||||||   |  | ||||
Db 191975 TCACAACAAACATAACAAACAAACTAGTTGACTCATTAAAACAAAATTTTAAGAAAATCT 192034

Qy      82 GACAGTGATTTAAATTTAAATGACGTAAAAGTTAGATATTATTACACAAGTGATGGTACA 141
             | |   || |    ||  || |  || || | | ||||  ||| | |  | |   || |
Db 192035 AAAATAAATATTTTATTTTATAAAATAGAATTAATATATATTTATAAATTTTAAATTAAA 192094


Search completed: February 22, 2003, 19:04:18
Job time : 324.722 secs
```